

THE
LONDON, EDINBURGH AND DUBLIN
PHILOSOPHICAL MAGAZINE
AND
JOURNAL OF SCIENCE

CONDUCTED BY
SIR LAWRENCE BRAGG, O.B.E., M.A., D.Sc., F.R.S.
SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.
AND
ALLAN FERGUSON, M.A., D.Sc.

"Nec araneorum sane textus ideo melior quia ex se fila gignunt, nec
noster vilior quia ex allenis libamus ut apes."

JUST. LIPS. *Polit. lib. i. cap. l. Not.*

VOL. XXXV.—SEVENTH SERIES
JANUARY—DECEMBER 1944

LONDON:
TAYLOR & FRANCIS, LTD., RED LION COURT, FLEET STREET, E.C. 4.
Agents: WM. DAWSON & SONS, LTD., 129 CANNON STREET, E.C. 4.

"Meditationis est perscrutari occulta; contemplationis est admirari perspicua Admiratio generat quæstionem, quæstio investigationem, investigatio inventionem."—*Hugo de S. Victore.*

— "Cur spirent venti, cur terra dehiscat,
Cur mare turgescat, pelago cur tantus amaror,
Cur caput obscura Phœbus ferrugine condât,
Quid toties diros cogat flagrare cometas,
Quid pariat nubes, veniant cur fulmina cœlo,
Quo micet igne Iris, superos quis conciat orbes
Tam vario motu."

J. B. Pinelli ad Mazonium.



CONTENTS OF VOL. XXXV. (Nos. 240-251)

(SEVENTH SERIES)

NUMBER 240—JANUARY

	Page
I. The New Algebras and their Significance for Physics and Philosophy. By E. T. WHITTAKER	1
II. Note on Tables of an Integral. By ALAN FLETCHER, M.A., Ph.D.	16
III. The Effect of Shape on the Thermal Endurance of Glass Rods. By J. B. MURGATROYD, B.A., F.S.G.T.	17
IV. Thermal Diffusion in Mixtures of Molecules of Small Mass Difference. By K. E. GREW, Ph.D.	30
V. Effect of Space Charge on Electron Beams. By J. AHARONI, Department of Physics, King's College, London	36
VI. Note on the Solubility of Hydrogen in Palladium. By Professor E. A. OWEN, M.A., Sc.D., University College of North Wales, Bangor....	50
VII. Viscosities at the Boiling Point of some Primary Amines, Cyclo- hexane and some of its Derivatives. By J. NEWTON FRIEND and WILLIAM D. HARGREAVES	57
VIII. Modifications of Appleton's Method of Measuring the Mutual Conductance of a Valve. By L. F. BATES, Professor of Physics, and W. F. LOVERING, University College, Nottingham	64

NUMBER 241—FEBRUARY

IX. Molecular Weight, Gram Molecule and Avogadro's Number. By W. E. DUNCANSON, M.Sc., Ph.D.	73
X. Terminology and Symbols in Physics. By W. E. DUNCANSON, M.Sc., Ph.D.	81
XI. The Geometry of Demagnetizing Curves, and its Consequences for Permanent Magnet Materials. By K. HOSELITZ, Ph.D.	91
XII. On a Problem in Heat Conduction. By C. J. TRANTER, M.A.	102
XIII. Introductory Notes on a Reformulation of the Special Theory of Relativity. By P. A. SCOTT-IVERSEN, F.R.A.S.	105
XIV. New Contributions to Interferometry. Part II.—New Interfer- ence Phenomena with Newton's Rings. By S. TOLANSKY, Ph.D., D.I.C., The University, Manchester. (Plates I.-III.)	120
XV. Viscosities of Unsaturated Six-Membered Isocyclic Compounds. By J. NEWTON FRIEND and WILLIAM D. HARGREAVES	136

NUMBER 242—MARCH

XVI. Transient Response in Frequency Modulation. By D. A. BELL, M.A., B.Sc., A.M.I.E.E.	143
XVII. Graphical Computation. By C. W. HANSEL, Bedford School ..	159
XVIII. Some Problems involving Line Sources in Conduction of Heat . By J. C. JAEGER	169

	Page
XIX. New Contributions to Interferometry. Part III.—The Differential Polarization Phase Change on Reflection at a Thin Silver Film. By S. TOLANSKY, Ph.D., D.I.C., The University, Manchester.....	179
XX. Remarks on Electrical Units. By D. P. DALZELL.....	188
XXI. The Ultrasonic Diffraction Grating. By R. A. HOUSTOUN, M.A., D.Sc., University of Glasgow.....	192
XXII. A Non-recording Microphotometer with Variable Contrast Sensitivity. By T. B. RYMER, Ph.D., and C. C. BUTLER, B.Sc., The University, Reading	202
XXIII. The Development of the Trichromatic Theory of Colour Vision. By Prof. W. PEDDIE	209
XXIV. New Contributions to Interferometry, Part I.....	210
XXV. New Contributions to Interferometry.....	211
XXVI. New Contributions to Interferometry, Part I. A Reply	211
XXVII. Notices respecting New Books :—	
Mr. J. E. CALTHROP'S A German Physics Reader.....	213
Reports on Progress—Voics. in Physl. IX. (1942–1943)	213

NUMBER 243—APRIL

XXVIII. The Influence of Crystal Size on the Absorption Factor as applied to Debye-Scherrer Diffraction Patterns. By A. TAYLOR, M.Sc., Ph.D., F.Inst.P.	215
XXIX. New Contributions to Interferometry. Part IV.—The Ghost Images and Scatter Rings of the Fabry-Perot Interferometer and their Effects in Hyperfine Structure Observations. By S. TOLANSKY, Ph.D., D.I.C., Manchester University. (Plate IV.)	229
XXX. Kinematic Relativity. By Professor W. WILSON, F.R.S.....	241
XXXI. Solutions of some Problems in Viscous Flow. By GEORGE GREEN, D.Sc., Lecturer in Applied Physics in the University of Glasgow.	250
XXXII. Coefficients for Numerical Integration with Central Differences. By HERBERT E. SALZER.....	262
XXXIII. Equilibrium of a Thin Slab on an Elastic Foundation of Finite Depth. By A. H. A. HOGG	265
XXXIV. On Under-expansion in a Steam-nozzle, with Reference to an Unusual Case. By L. J. KASTNER, M.A.	276

NUMBER 244—MAY

XXXV. Methods for the Investigation of the Properties of Liquids at Ultra High Frequencies. By GORDON WILLIAMS, M.Sc., Ph.D., F.Inst.P., Wheatstone Laboratory, King's College, London	283
XXXVI. On the Dimensions of Physical Magnitudes (Fourth Paper). By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.	296
XXXVII. Problems of the Transmission of Light through Tissues and some other Media. By DAVID S. EVANS, M.A., Ph.D.....	300
XXXVIII. A new Approach to the Dynamics of Systems with Gyroscopic Coupling Terms. By A. BLOCH, Dr.-Ing., M.Sc. (Communication from the Staff of the Research Laboratories of The General Electric Company, Limited, England.)	315

	Page
XXXIX. A Centrifugal Method of Measuring the Surface Tensions and Interfacial Tensions of Liquids in Capillary Tubes. By W. MEYERSTEIN, M.D. (Department of Physiology in the University of Birmingham), and J. D. MORGAN, D.Sc.	335
XL. An Investigation of the Existence of "Magnetic" Currents. By D. M. MILLEST, B.Sc.	342
XLI. On Helmholtz's Electrodynamical Potential and the Induction Coefficients of Unclosed Currents in Quasi-stationary Processes. By A. DA SILVEIRA, Professor of Physics at Instituto Superior Tecnico, Lisbon. .	346
XLII. Notices respecting New Books :—	
Sir AMBROSE FLEMING : (1) Mathematics for Engineers ; (2) Physics for Engineers ; (3) Chemistry for Engineers ; (4) Mechanics for Engineers	351
Prof. P. W. SELWOOD's Magnetochemistry	353
Tables of the Bessel Functions $J_0(z)$ and $J_1(z)$ for Complex Arguments	353
Table of the Reciprocals of the Integers from 100,000 through 200,000	354

NUMBER 245—JUNE

XLIII. Two Methods of Measuring Ultra High Frequency Electric Fields. By K. RACHAEL MAKINSON, B.A. (Cantab.), A.Inst.P., Commonwealth Research Scholar in the University of Sydney (1940-41), and H. D. FRASER, B.Sc. (Sydney).....	355
XLIV. The Reception of Ohm's Electrical Researches by his Contemporaries. By H. J. J. WINTER M.Sc., A.Inst.P., University College of the South-West of England, Exeter.....	371
XLV. A Note on Two Definitions of Noise Figure in Radio Receivers. By D. K. C. MACDONALD, M.A., A.Inst.P., Captain, R.E.M.E., Instructor, Military College of Science, Fire Control School	386
XLVI. The Accumulation of Chance Effects and the Gaussian Frequency Distribution. By LUDWIK SILBERSTEIN, Ph.D.	395
XLVII. On the Application of the Micro-absorption Factor to Problems of Lattice Distortion and the Nature of Anti-phase Domains. By A. TAYLOR, M.Sc., Ph.D., F.Inst.P.	404
XLVIII. Exercises on a Tightrope. By E. H. NEVILLE	414
XLIX. Dimensions. By Professor W. WILSON F.R.S.	420
L. Notices respecting New Books :—	
Dr. G. W. SCOTT-BLAIR's Survey of General and Applied Rheology	426
Prof. MAX BORN's Experiment and Theory in Physics.....	426

NUMBER 246—JULY

LI. A General Theory of Antennæ. By J. AHARONI, Ph.D.....	427
LII. Further Properties of Recurrent Exponential and Probability Function Pulse Wave-forms. By F. F. ROBERTS, B.Sc. (Eng.) and J. C. SIMMONDS, Ph.D.	459
LIIL. Conditions of Failure and Factors of Safety in Members subject to Alternating Stresses. By S. V. SERENSEN, Member Acad. Sci. of the Ukraine	470

	Page
LIV. Experiments on the Six Focal Lines due to Reflexions at the Surfaces of Two Plano-Cylindrical Lenses. By G. F. C. SEARLE, Sc.D., F.R.S., Cavendish Laboratory, Cambridge	477
LV. On the Measurement of the Residual Parameters of a "Q" Meter. By W. F. LOVERING	491
LVI. Note on Green's Functions in the Theory of Heat Conduction. By ARNOLD N. LOWAN	495
LVII. Notices respecting New Books :— Messrs. G. H. HARDY and W. W. ROGOSINSKI'S Fourier Series	498

NUMBER 247—AUGUST

LVIII. A Relativistic Theory of Temperature Radiation. By HERBERT DINGLE, Professor of Natural Philosophy, Imperial College of Science and Technology	499
LIX. Solutions of Problems Relating to Media in Contact by the Method of Wave-trains. By GEORGE GREEN, D.Sc., Lecturer in Applied Physics, The University, Glasgow	519
LX. Note on Certain Integrals Involving Hermite's Polynomials. By T. D. H. BABER and LEONID MIRSKY	532
LXI. Solution of the Restricted Problem of the Random Walk. By LUDWIK SILBERSTEIN, Ph.D.	538
LXII. The Measurement of X-ray Wave-lengths by the Powder Method : $\text{CrK}\beta_1$ and $\text{MnK}\beta_1$. By H. LIPSON and LILLIAN E. R. ROGERS, Cavendish Laboratory, Cambridge	544
LXIII. Two-Dimensional Compressible Flow past a Solid Body in Unlimited Fluid or Symmetrically Placed in a Channel. By S. GOLDSTEIN, F.R.S., and M. J. LIGHTHILL, B.A., of the Aerodynamics Department, National Physical Laboratory	549
LXIV. Notices respecting New Books :— Sir WILLIAM CECIL DAMPIER'S Shorter History of Science Prof. A. A. LUCE'S Berkeley's Philosophical Commentaries.	568 569

NUMBER 248—SEPTEMBER

LXV. On the Tones of Thin Cylindrical Tubes. By GEORGE E. ALLAN, D.Sc.	571
LXVI. Eigen-Values and Eigen-Functions for the Operator $d^2/dx^2 - x $. By R. P. BELL, F.R.S. (Balliol College, Oxford)	582
LXVII. The Development of an Automatic Electric Balance for Research on Aerodynamic Stability. By J. B. BRATT, B.A., B.Sc., and K. C. WIGHT, of the Aerodynamics Division, N.P.L.	588
LXVIII. Influence of Air-gases on the First and Second Positive Systems of Nitrogen. By N. R. TAWDE, M.Sc. (Bom.), Ph.D. (Lond.), F.Inst.P., F.N.I., and V. S. PATANKAR, M.Sc., Ph.D. (Bom.), Royal Institute of Science, Bombay	600
LXIX. An Electrical Circuit for Harmonic Analysis and other Calculations. By T. B. RYMER, Ph.D., and C. C. BUTLER, B.Sc., The University, Reading	606
LXX. On the Dimensions of Physical Magnitudes (Fifth Paper). By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.	611

	Page
LXXI. Viscosities and Rheochors of Aldehydes, Nitriles and of Secondary and Tertiary Amines. By J. NEWTON FRIEND and WILLIAM D. HARGREAVES	619
LXXII. On the Optimum Thickness of Powder Specimens in X-ray Diffraction Work. By A. TAYLOR, M.Sc., Ph.D., F.Inst.P.	632
LXXIII. The Effect of Heterogeneity in Powders on the Intensity of X-ray Reflections. By G. W. BRINDLEY	638
LXXIV. Reply to Dr. Brindley's letter on "The Effect of Heterogeneity in Powders on the Intensity of X-ray Reflections." By A. TAYLOR	640
LXXV. Notice respecting New Book :— DR. SANDELL'S Colorimetric Methods for the Determination of Traces of Metals	641

NUMBER 249—OCTOBER

LXXVI. A New Photo-Electric Method for Fourier Synthesis and Analysis. By R. FÜRTH and R. W. PRINGLE, Edinburgh	643
LXXVII. A Note on a Certain Multiple Integral. By A. S. GLADWIN, B.Sc., A.R.T.C., A.M.I.E.E. (Communication from the Staff of the Research Laboratories of The General Electric Co., Ltd., England.)	657
LXXVIII. Some Devices for the Solution of Large Sets of Simultaneous Linear Equations. (With an Appendix on the Reciprocity of Partitioned Matrices.) By Prof. W. J. DUNCAN, D.Sc., M.I.Mech.E.	660
LXXIX. Expansions in Terms of Associated Legendre Functions. By T. M. MACROBERT (Glasgow)	670
LXXX. Temperature, Pressure, and Specific Volume Changes of a Gas Mixture under Dissociation and Re-association Conditions, with Combustion following the Straight Line Law, $P=mV+n$. By W. J. WALKER, D.Sc., Ph.D., Professor of Mechanical Engineering, University of the Witwatersrand, Johannesburg, South Africa	680
LXXXI. Sums of Series of the Form $\sum_{n=0}^{\infty} a_n J_{n+\alpha}(z) J_{n+\beta}(z)$. By S. O. RICE.	686
LXXXII. On the Temperature Coefficient of Air-cored Self-inductances. By A. BLOCH, Dr.Ing., M.Sc. (Communication from the Staff of the Research Laboratories of The General Electric Company, Limited, England.)	693
LXXXIII. On the Theory of Seismometers. By C. G. PENDSE, M.A., Ph.D.	706
LXXXIV. Notices respecting New Books :— Tables of Lagrangian Interpolation Coefficients prepared by the Mathematical Tables Project	713
Table of Circular and Hyperbolic Tangents and Co- tangents for Radian Arguments, by the Mathematical Tables Project	714

NUMBER 250—NOVEMBER

LXXXV. A Push-Pull Resistance-Capacity Coupled Oscillator. By W. F. LOVERING, B.Sc.	715
LXXXVI. The "Escalator" Process for the Solution of Lagrangian Frequency Equations. By J. MORRIS, B.A., and J. W. HEAD, M.A.	735

	Page
LXXXVII. Thermal Transpiration and Reversible Process for a Degenerate Gas. By D. V. GOGATE, M.Sc., Ph.D. (London) and V. N. UPADHYAYA, B.Sc., Physics Department, Baroda College, Baroda (India).	760
LXXXVIII. An Analysis of Hardness. By A. C. VIVIAN, D.Sc.	765
LXXXIX. The Physical Realizability of Electrical Networks having Prescribed Characteristics, with particular reference to those of the Probability Function Type. By F. F. ROBERTS, B.Sc. (Eng.) and J. C. SIMMONDS, Ph.D.	778
XC. A Note on the Minimum Radius for Degenerate Stellar Masses. By D. S. KOTHARI and F. C. AULUCK (University of Delhi).	783
XCI. Notice respecting New Book :— Mr. G. W. CARTER's The Simple Calculation of Electrical Transients.....	786

NUMBER 251—DECEMBER

XCII. Energy Distribution in the Spectrum of a Frequency Modulated Wave.—Part I. By A. S. GLADWIN, B.Sc., A.R.T.C., A.M.I.E.E.	787
XCIII. Diffusion in Spherical Shells, and a New Method of Measuring the Thermal Diffusivity Constant. By R. M. BARRER, D.Sc., Ph.D., F.R.I.C.	802
XCIV. The tabulation of some Bessel Functions $K_\nu(x)$ and $K_\nu'(x)$ of Fractional Order. By H. R. F. CARSTEN, Dr.-Ing., F.Inst.P., and Miss N. W. McKERROW, B.Sc.	812
XCV. The Field between Equal Semi-infinite Rectangular Electrodes or Magnetic Pole-pieces. By N. DAVY, University College, Nottingham..	819
XCVI. Alternating Loads on Sleeve-Bearings. By J. DICK, B.Sc., Ph.D.	841
XCVII. An Application of the Method of Finite Difference Equations to a Problem of Bending Moments. (Continuous Beam of N Equal Spans under Uniform Loading.) By N. J. DURANT	848
XCVIII. Letter from Mr. P. K. Chatterjea and reply by Mr. D. A. Bell to the paper on "Transient Response in Frequency Modulation"	851
XCIX. Letter from Mr. H. E. Curtis and reply by Mr. D. A. Bell to the paper on "Transient Response in Frequency Modulation".....	852
C. Notice respecting New Book :— MR. DOUGLAS W. HILL's The Impact and Value of Science	854
Index.....	855

P L A T E S

- I.-III. Illustrative of Dr. S. TOLANSKY's Paper on New Contributions to Interferometry. Part II.—New Interference Phenomena with Newton's Rings.
- IV. Illustrative of Dr. S. TOLANSKY's Paper on New Contributions to Interferometry. Part IV.—The Ghost Images and Scatter Rings of the Fabry-Perot Interferometer and their Effects in Hypersine Structure Observations.

THE
LONDON, EDINBURGH & DUBLIN
PHILOSOPHICAL MAGAZINE
AND
JOURNAL OF SCIENCE

[SEVENTH SERIES—VOL. 35]

I. *The New Algebras and their Significance for Physics and Philosophy.*

By E. T. WHITTAKER*.

(Address of the President at the Annual Statutory Meeting of the
Royal Society of Edinburgh, October 25th, 1943).

[Received December 1, 1943.]

It was a hundred years ago this month—to be precise, on the 16th October, 1843—that Hamilton discovered quaternions. The event was of major importance, for the quaternion calculus was the prototype of all the symbolism which disobey the rules of ordinary algebra, and which are now so remarkably useful in the philosophical and physical applications of mathematics: such as the Boolean and other systems of symbolic logic, and the matrix theory which is the operative method of quantum mechanics.

In our own Society there are special reasons for observing the centenary, for after Hamilton's death his mantle fell on his disciple Tait, and in the second half of the nineteenth century Edinburgh was the stronghold of the quaternionists. It seems appropriate therefore to devote the address on this occasion to a consideration of the origin and development of the new algebras, with special reference to their value for the investigation of the external world.

The pupil who begins the study of ordinary algebra is told that the

* Communicated by the Author.

letters which occur in it, a, b, c, d, x , and so forth, represent *numbers*; and the laws according to which algebraic operations are conducted are laws which are actually obeyed by numbers: thus the product obtained when we multiply 2 by 3 is equal to the product obtained when we multiply 3 by 2; and this *commutative law of multiplication*, as it is called, is embodied in the formula

$$ab=ba.$$

From the third century, when Diophantus of Alexandria invented the beginnings of an algebraical notation, to 1843, this law was undisputed, as indeed it must be so long as algebra is merely symbolized arithmetic. The possibility that an algebraical letter might stand for something other than a real number was first envisaged more than two hundred years ago, when the so-called "imaginary unit," $\sqrt{-1}$, was introduced and its properties were investigated. Such a formula as

$$\pi = \frac{\log(-1)}{\sqrt{-1}},$$

in which the ratio of the circumference of a circle to its diameter is expressed in terms of this new quantity, stirred the imagination of the eighteenth-century mathematicians. But $\sqrt{-1}$, though not a number in the old sense, obeyed the commutative law of multiplication and the other classical rules, and was essentially a new arithmetic entity, not an element of a new algebra; and another century was to elapse before Hamilton made his great adventure of breaking away from the rules themselves.

Let us enlarge our notions of mathematical symbolism by supposing that letters such as a, b, x, \dots can denote, not numbers but operations of some kind. The first question we have to ask is, What meaning should be attached to the product ab when a and b denote operations? Now in ordinary algebra, if we want to multiply a number x by the product of two numbers a and b , we can first multiply x by b , and then multiply the resulting number by a ; and analogously when a and b are operators, the product ab is most naturally defined to be the operator whose performance is equivalent to operating first with the operator b and then operating on the result with the operator a . This being so, now suppose that a man is facing north, and that the operation of making a right-about turn so as to face south is denoted by a , while the operation of raising the arm which is on his west side is denoted by b ; then in the compound operation ab he first raises his west arm, that is, his left arm, and then turns round, so that finally he is facing south with his left arm raised. On the other hand, the product ba would mean that he performs the operation a first, that is, he turns round, and then performs the operation b , which is raising his west arm, that is, his right arm, so finally he is facing south with his right arm raised. Thus the

result of the operation ab is different from the result of the operation ba , and the commutative law

$$ba=ab$$

is not valid in this symbolism.

Hamilton's quaternions were devised to describe geometrical operations such as rotations, and therefore we need not be surprised to learn that the commutative law does not hold in this calculus. But although this may now seem to us quite natural, and indeed almost obvious, the surrender of the law was a tremendous break with tradition when it was first introduced, and marked the beginning of a new era. As soon as the shock of the innovation had passed, quite a number of non-classical algebras were introduced. I will speak first of one which initiated important new developments in logic, and, through logic, in the philosophy of mathematics.

As every mathematician knows, no rigorous thinking can be done in ordinary language. A word or a phrase usually has many different meanings, and in verbal discussions it is practically impossible to avoid occasions when one meaning is unconsciously substituted for another, or a tacit assumption is surreptitiously introduced. Long ago Leibnitz realized that in order to make valid reasoning possible on any difficult subject it would be necessary to construct an ideography like that of algebra, capable of doing what ordinary language cannot do, that is, to represent ideas and their connections without introducing undetected assumptions and ambiguities. He therefore conceived the idea of a logical calculus, in which the elementary operations of the process of reasoning would be represented by symbols—an alphabet of thought, so to speak—and envisaged a distant future when philosophical and theological discussions would be conducted by its means and would reach conclusions as incontrovertible as those of mathematics. The project, however, made no appreciable progress until, under the stimulus of Hamilton's modification of the laws of algebra, it was taken up by George Boole, for the latter part of his life Professor in Cork, who published a sketch of his theory in 1847, four years after the discovery of quaternions, and a fuller account in 1854 in his book '*An Investigation of the Laws of Thought*.'

Boole's purpose was to symbolize logic: that is, to express the fundamental laws of reasoning in the language of a calculus, so as to exhibit logic as a system of processes carried on by the aid of symbols, and subject to rules of transformation founded upon the meaning of these symbols alone. In this system, a letter such as x denotes a *class* or collection of individual things to which some common name can be applied: for instance, x might represent the class of all doctors. We can also regard x as a symbol of operation, namely, the operation which selects, from the totality of objects in the world, or in the "universe of discourse," those objects which are doctors. Now let y denote some

other class, say the class of all women. Then, according to the principles we have already laid down, the product xy must represent the result of first selecting all women, and then selecting from them those who are doctors: that is, xy represents all women doctors, all the individuals who belong both to the class x and to the class y . Evidently the order in which the two symbols stand is indifferent, so

$$xy=yx;$$

in the new algebra the commutative law holds for multiplication. When, in ordinary language, a noun is qualified by an adjective, as in "feminine doctor," we must understand the idea represented by this product.

Now consider the case when the class y is the same as the class x . In this case, the combination xy expresses no more than either of the symbols taken alone would do, so $xy=x$, or (since y is the same as x)

$$x^2=x.$$

In ordinary algebra, the equation $x^2=x$ is true when x has either of the values zero and unity, but in Boolean algebra all symbols obey this law.

Let us next take up the question of addition. Boole defined $x+y$ to be the class consisting of all the individuals who belong either to the class x or to the class y : he supposed the classes to have no members in common, but this restriction was abolished by his successors, so we now say that the class $x+y$ consists of all the individuals who belong to *one at least* of the classes x and y , whether the classes overlap or not. Evidently we shall have

$$x+y=y+x,$$

so the commutative law holds for addition.

The symbol used for zero in ordinary algebra is used in Boolean algebra to denote the class which has no members, the null class: obviously we must have

$$x.0=0 \quad \text{and} \quad x+0=x,$$

as in ordinary algebra.

The symbol 1 is used to denote the class consisting of everything, or the universe of discourse; it has the properties

$$x.1=x \quad \text{and} \quad x+1=1.$$

Lastly, the minus sign must be introduced: the symbol $-x$ is defined to be the class consisting of those members of 1 which do not belong to x ; so that

$$-x+x=1 \quad \text{and} \quad x.(-x)=0.$$

So far, we have interpreted Boolean algebra as an algebra of classes; but we may take the classes to be classes of cases in which certain propositions are true, and this led to an interpretation of it as an algebra

of propositions; if x and y are propositions, their product would represent simultaneous affirmation, so xy would be the proposition which asserts "both x and y ": the sum would denote *alternative affirmation*, so $x+y$ would be the proposition "either x or y or both." The minus sign would represent "it is not true that," so $-x$ would be the proposition contrary to x ; the equation $x=1$ would imply that x is true, while the equation $-x=1$, which is equivalent to $x=0$, would signify that x is false. The equation $-x+x=1$ would now represent the logical Principle of the Excluded Middle, and the equation $x.(-x)=0$ would represent the Principle of Contradiction.

One of the elementary processes of logic consists in deducing from two propositions, containing a common element or middle term, a conclusion connecting the two remaining terms. This corresponds to the process of elimination in algebra, and may be performed in a way roughly analogous to it. The parallelism of logic and algebra is indeed far-reaching; for instance, the logical distinction between categorical propositions and conditional propositions corresponds closely to the algebraical distinction between identities and equations. Again, the inequalities of algebra have their analogues in logic. Consider, for instance, the statement that if a proposition a implies a proposition b , and b implies a proposition c , then a implies c . This bears an obvious resemblance to the algebraical theorem that if a is less than b , and b is less than c , then a is less than c . It is useful to have a symbol which represents logical implication or inclusion, and all modern forms of symbolic logic do in fact employ one or two, one in the calculus of propositions and one in the calculus of classes; this, however, does not represent an independent concept, but can be defined in terms of the logical product; for the statement that a is included in, or implies, b is equivalent to the statement that the logical product of a and b is equal to a .

When the traditional logic was examined by the methods of the new symbolism, many errors and imperfections were discovered. For example, of the nineteen moods of the syllogism which were catalogued in the mnemonic *Barbara Celarent*, four were revealed as false, while the remaining fifteen were seen to be reducible to one of them (the syllogism in *Barbara*) by very simple transformations, so that their discussion separately was unnecessary. The most important advance, however, was not the detection of the errors of the old logic, but the removal of its limitations. The Aristotelian system in effect took into account only subject-predicate types of propositions, and failed to deal satisfactorily with reasoning in which relations were involved, such as "If there is a descendant, there must be an ancestor." In this and other respects the subject has become enlarged to such an extent that only a comparatively small part of any modern treatise is devoted to the traditional logic.

At the beginning of this lecture I spoke of Hamilton's quaternions as

the prototype of developments in symbolic logic. From what has been said so far, it may perhaps have seemed as if the relationship amounted to nothing more than this, that Hamilton was the first to make a breach in the structure of classical algebra, and that Boole and his successors then assaulted it at other points and in radically different ways. Certainly the rules of Boolean algebra have little resemblance to the quaternionic formulæ, and the most striking novelty of Hamilton's calculus, the non-commutative law of multiplication, was not adopted by any of the earlier mathematical logicians. However, a much closer connection with quaternions was shown in the 'Logic of Relatives,' which was published by the American philosopher Charles S. Peirce in 1870. To take a simple example, consider two classes such that every individual of one class bears a certain relation to every individual of the other class: say, let the two classes be the masters and the boys in a school where every master teaches every boy. Then there are four elementary relatives (as Peirce called them) involved, namely, master to boy (teacher), boy to master (pupil), master to master (colleague), and boy to boy (schoolmate). Moreover, there are sixteen propositions, of which the following are samples:—

The pupils of the teachers of any person are that person's schoolmates.

The colleagues of the colleagues of any person are that person's colleagues.

There are no pupils of any person's schoolmates.

These may be regarded as the "multiplication-properties" of the relatives, and may be arranged in a table thus (where *c*, *t*, *p*, *s* stand for colleague, teacher, pupil, schoolmate):

	<i>c</i>	<i>t</i>	<i>p</i>	<i>s</i>
<i>c</i>	<i>c</i>	<i>t</i>	o	o
<i>t</i>	o	o	<i>c</i>	<i>t</i>
<i>p</i>	<i>p</i>	<i>s</i>	o	o
<i>s</i>	o	o	<i>p</i>	<i>s</i>

The multiplier is in the left-hand column and the multiplicand in the top row, the product being found in the corresponding compartment in the middle.

This table presents an obvious analogy to the multiplication table for the quaternion units $1, i, j, k$, which is as follows :—

	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1

And in fact the algebra based on the former multiplication table can be reduced to that based on the latter.

Now any theorem in the geometry of three-dimensional space can be represented by a formula in quaternions ; and, as we have just seen, quaternionic formulæ are essentially equivalent to formulæ in Peirce's calculus of relatives. Thus we arrive finally at the remarkable conclusion that *to every proposition in geometry there corresponds a proposition in the logic of relations.*

The advantages of an ideography as compared with ordinary language are strikingly evident in the discussion of certain contradictions which have threatened to invalidate reasoning, such as the famous paradox which was discovered forty years ago by Bertrand Russell. He remarked that in the case of, *e. g.*, the class whose members are all thinkable concepts, the class, being itself a thinkable concept, is one of its own members. This is not the case with, *e. g.*, the class of all blue objects, since this class is not itself blue. We can therefore say that those classes which do not contain themselves as one of their members form a particular kind of classes : the aggregate of these classes constitutes a new class which we shall call x . Let us put this definition in the two forms :

Form A. A class which contains itself as a member is not a member of x .

Form B. A class which does not contain itself as a member is a member of x .

Now if x were a member of itself, then by *A* it would not be a member of itself, so we should have a contradiction ; while if x were not a member of itself, then by *B* it would be a member of itself, which is again a contradiction. Thus on either supposition we arrive at a contradiction, which appears to be insoluble by any kind of verbal explanation.

Now let us look at the matter from the point of view of symbolism. The contradiction that " x is an x " is equivalent to " x is not an x " was obtained essentially by substituting x for y in the statement that (1) y is a class, (2) y is an x , is equivalent to " y is not a y ." This substitution, however, is not, as it stands, an operation performed on the fundamental logical symbols in accordance with the rules which were laid down for operating on them: for x is not itself one of the elementary ideograms, but is an abbreviation, a single letter standing proxy for a complex of ideas. Now all abbreviations, however convenient, are from the logical point of view superfluous; and an argument involving them is not valid unless, at every stage of it, the proxy symbols can be replaced by the full expressions for which they stand. In order therefore to be sure that what has been done is correct from the point of view of symbolic logic, we must translate the whole argument, and in particular the operation of substituting x for y , into the language of the elementary ideograms and the operations that are permissible with them, so that all explicit mention of x will have been eliminated from the proofs. When, however, we try to do this, we find that we cannot. *It is not possible to state Russell's paradox in the form of an assertion composed solely of the elementary ideograms.* This shows that if we had from the beginning avoided the use of ordinary speech or of proxy symbols, and conducted all our investigations according to the strict precepts of ideography, then Russell's paradox would never have emerged. It can be obtained by argumentation in words, or it can be obtained by a quasi-symbolic argument in which an operation is permitted which is untranslatable into pure symbolic logic; but it cannot be obtained by any process which is restricted to using throughout nothing but the elementary ideograms and the operations that are recognized as permissible with them, and which expresses both the final result and all intermediate equations in terms of them exclusively. Thus Russell's paradox, being inexpressible in symbolic logic, is really meaningless, and we need not concern ourselves with it further: the contradiction which appears in it is not inherent in logic, but originates in the imperfections of language and of abbreviated symbolism.

When Boole and his successors replaced the verbal reasonings of the old logic by operations conducted in the mathematical fashion with ideographic symbols, it was natural that the movement should react on mathematics itself. The mathematicians, accepting Benjamin Peirce's definition of mathematics as "the science that draws necessary conclusions," recognized its relation to the science of *drawing* necessary conclusions, which is logic, and perceived in the new symbolism a powerful instrument for investigating the fundamental concepts of their own science, such as *number* and *infinity*. The doctrine regarding the relation of the two subjects which is maintained by many writers, and particularly by Whitehead and Russell, is that mathematics is simply a subdivision of logic: a subdivision which has undergone an exceptionally great development as compared with general logic, partly because there

are economic motives for its cultivation, and partly because its progress has been facilitated by the invention of a powerful symbolism. Whitehead and Russell have expounded this thesis in their great work '*Principia Mathematica*,' and have defined the whole numbers one, two, three, . . . , on which mathematics is based, in terms of purely logical concepts such as *class*, *implication*, *negation*, Whether their deduction is free from imperfections has been disputed; but this may at any rate be said, that if the verdict should go against them, and it should be established that mathematics contains elements which do not belong to logic in its present form, then it would still be possible to enlarge the logical symbolism by the adjunction of new elements until it is capable of providing a complete foundation for mathematical science.

It is, then, to the new algebras that the logician and the pure mathematician look for the attainment of complete rigour and accuracy in their formulations. Very far removed from this is the attitude of the mathematical physicist, who is interested in non-commutative symbolism for a wholly different reason. He may be said to be, in a certain sense, moving *away* from precision; for whereas classical physics was based on the notion of the particle, which, mathematically speaking, is a moving geometrical point and is quite precise, modern physics, on the other hand, has shown that the elementary constituents of the universe—electrons, protons, etc.—are not simply particles, but have wave properties also. The resulting indefiniteness or imperfect definition cannot be adequately described in terms of the concept of the moving geometrical point, and calls for the introduction of a new algorithm, specially fitted to deal with it. It has been one of the most remarkable discoveries of the present century that such an algorithm is actually provided by one of the varieties of non-commutative algebra.

It may be remarked that the concept of the geometrical point had for some time been subjected to criticism by the mathematical philosophers, quite independently of the situation created by discoveries in atomic physics. The Euclidean point is, in fact, not an object of immediate perception, but is, in its logical aspect, a somewhat remote and complex entity, arrived at only by an endless process of abstraction: what we actually meet with in the external world are not points, but volumes. Starting from this idea, attempts have been made to create a theory of geometry in which the volume rather than the point is taken as the fundamental element; instead of regarding a volume as a class of points, the point is derived as the limit of a sequence of volumes, and the primitive or undefined notions assumed at the outset are such as apply to volumes, *e. g.*, the relation of one volume to another which entirely includes it. Unfortunately such a geometry, while closer to actual experience than a geometry based on points, is much more complicated mathematically; and indeed it is not until the point has been defined, and has been established as the dominating concept, that theorems of any geometrical interest are obtained.

For the purposes of atomic physics, however, this is not the kind of

geometry that is wanted: the blurring or indefiniteness which we have to deal with in quantum theory is not exactly analogous to the indefiniteness of a location regarding which we know only that it is somewhere within a certain volume. To explain its nature, an analogy from music may be useful.

Let us imagine an experimental arrangement in which a musical note is produced by an organ-pipe, controlled by a key, so that the sound is emitted when the key is pressed down, and is cut off instantly when the key is released. The note will be supposed to be a pure tone, *i. e.* a simple harmonic oscillation of the air, and to be of frequency (that is, rapidity of vibration) ν , very low in the scale, so that the number of oscillations per second is comparatively small. Then when the experiment is performed, the question "At what instant was the note of frequency ν sounded?" does not admit of a precise answer, since the sound would actually extend over the finite interval of time while the key was down. In order to obtain as nearly as possible a note sounded at a precise instant, let us shorten its duration as much as possible, releasing the key a fraction of a second after it has been depressed. In this way we can increase the precision of the statement regarding the time of the sound; but, by doing so, we cut short the train of oscillations; and if we imagine the duration of the note shortened so much that less than a single complete oscillation takes place, the sound will no longer produce on the ear the sensation of a definite pitch: it can no longer be described as of frequency ν .

Thus the two requirements, that the sound should be emitted at a definite instant and that it should be of a definite pitch, are incompatible with each other: the more nearly instantaneous we can make the note, the less pure is its tone.

This inability to specify exactly and simultaneously all the quantities involved is characteristic of all wave-phenomena; if we wish to determine the rapidity of vibration, or frequency, of the motion, then it is necessary to observe the succession of undulations over a finite interval of time, and it is therefore impossible to specify any one instant as the "instant at which the measurement was obtained."

Similar insuperable obstacles to precise statement are found everywhere in atomic physics, and are the cause of the hazy or blurred definition which has already been mentioned as characteristic of elementary effects in nature. It may be shown that physical quantities generally are associated in pairs, such that an increase of accuracy in knowledge regarding one member of a pair necessarily entails a decrease of accuracy in knowledge regarding the other member.

As an illustration of this principle, suppose that we try to determine as accurately as possible the position of a free electron. We can imagine this done by means of a microscope pointed on the electron, which will be regarded for the moment as a particle, located at a definite point of space. It is well known that the distance apart of two objects which

can just be distinguished from each other in a microscope is proportional (*inter alia*) to the wave-length of the light employed ; thus the uncertainty in the measurement of the position of the electron is proportional to the wave-length of the radiation by which it is illuminated, and in order to make the precision of the measurement as great as possible, we must use radiation of very short wave-length ; by taking the wave-length sufficiently small we can, theoretically at any rate, determine the position to any desired degree of accuracy.

But an electron illumined by radiation scatters it, and in this scattering the electron experiences a recoil—what is known as the *Compton effect*—so that its momentum is changed by the act of observing its position ; and the violence of the recoil is inversely proportional to the wave-length of the radiation, while its direction is, within certain limits, not predictable. Thus in the process of measuring the position of the electron, its component of momentum parallel to the focal plane of the microscope is altered by an uncertain amount ; and the more precisely the position is determined (by diminishing the wave-length of the illumination), the greater is the uncertainty in the momentum. We can know one of them as accurately as we wish, but not both at the same time. It seems, therefore, as if spatial position and momentum have no separate and independent existence ; and this raises a problem which may be indicated by an analogy.

Suppose that a child with a penny comes to an automatic machine which supplies chocolate when the penny is put in one slot, and sweets when the penny is put in the other slot. Since he has only one penny, he can get either chocolate or sweets, but not both ; from the fact that he can get either at will, is he justified in concluding that they are both present in the machine ? Not necessarily ; for it is possible to imagine that there is a kind of paste in the machine which is converted into chocolate by his inserting the penny into one slot, and into sweets by his inserting the penny into the other. If this latter explanation is correct, then it is possible to imagine the machine fitted with a number of other slots, such that by inserting the penny into one of them a confection is obtained which is intermediate in some proportion between chocolate and sweets. This is analogous to the situation which exists in atomic physics. If we consider one of the elementary entities—electrons, protons, and so forth—which are the ultimate constituents of nature, we can have an accurate knowledge of its position, combined with complete ignorance as to its momentum, or we can have an accurate knowledge of its momentum, combined with complete ignorance as to its position, or we can have a simultaneous partial knowledge of both ; but there is no experimental justification for assuming that the entity is a particle in the old sense, possessing simultaneously an exact position and an exact momentum. We have no right to postulate the existence of entities which lie beyond the knowledge actually obtainable by observation, and which have no part in the prediction of future effects.

Thus the classical concept of a particle must be discarded ; in its stead there has been introduced a new fundamental element in the description of the external world, which is called a *state*.

Now the method of theoretical physics is essentially to analyse a complicated situation into an aggregate of elementary situations, each of which is governed by some simple law. Thus in the ordinary Newtonian mechanics of a system of bodies the interactions of the bodies are analysed into forces between pairs of particles ; and indeed throughout classical theory the ultimate elementary bodies are generally conceived as particles, each occupying a particular point of space at a particular moment, so that the concepts by aid of which the resolution is effected are the concepts of space and time. The great discovery of the present century has been that in atomic physics this method of analysis is wrong ; the blurring or imperfect definition, which has been described, simply expresses the fact that the true elementary constituents of nature have not the properties characteristic of Newtonian particles. There are events which extend over more than one point of space and more than one instant of time and which yet are *ultimate* events ; that is, they cannot be analysed into anything simpler than themselves. These are precisely the "states" of quantum mechanics.

It was necessary therefore to devise a new way of analysing situations, and a suggestion to this end was furnished by another of Newton's discoveries, namely the resolution of ordinary white light into its spectral components, violet, indigo, blue, green, etc. This resolves a compound wave-disturbance into its constituent simple undulations, each of which has a definite frequency ; it is entirely different in character from any analysis which depends on the resolution of space into points and of time into instants, and it serves to give a general idea of the analysis of a physical situation into states. With each state there is associated a definite *frequency* ; and states can be superposed on each other in much the same way as vibrations of different pitch can coexist in a single musical sound. Moreover, the frequency of an elementary state determines how it varies with the time ; and thus the resolution of the situation into states leads to a possibility of predicting its future development.

The difference between the old and new conceptions is seen very clearly when we consider the simplest of all systems, namely an electron or other elementary entity moving freely in a straight line under no external forces. According to classical physics, if a given particle is free to move in a given straight line, its motion is completely determined when a single numerical quantity is specified, namely its initial velocity (or, what comes to the same thing, its initial momentum) : the motion consists in a mere transference of the particle along the given line, with constant velocity. The classical concept of a particle excludes the possibility of its having more than one velocity in the same direction at any instant.

In contradistinction to this, let us consider the corresponding problem—the motion of an elementary entity in a straight line under no external

forces—as it is conceived in quantum mechanics. The phenomenon is now constituted by the superposition of a number of “states.” Each state, taken alone, would represent a particle with a definite momentum but (in accordance with the uncertainty principle) a completely unknown location: that is to say, if in a simple state we made an observation to determine the momentum of the particle, we should obtain a value which could have been predicted beforehand; but if we made an observation to determine its position, we should have no reason to expect it to be in one place rather than another. By superposing an infinite number of such states, we can obtain a disturbance which at any instant is strongly localized in the neighbourhood of a particular point: that is to say, if we made an observation to determine the position of the particle, there would be strong reason for expecting to find it near this point; but if we made an observation to determine the momentum, there would no longer be any certainty of obtaining a particular value: any one of an infinite range of values might be obtained, although there might be reason to expect a value not differing greatly from some value which could be predicted.

In order to describe mathematically the way in which the actual physical situation is built up from the infinitely many possible states, an arbitrary function is required; this is in sharp contrast to the classical problem, where the motion is specified by a single parameter, the velocity or momentum.

The history of the system in time can be calculated from the knowledge that every elementary state behaves in a way determined by its own frequency, and that the complete phenomenon at any instant is the superposition of the states as they are at this instant. Since the different states have different frequencies, and so do not keep in step with each other, the situation as a whole is not merely moving forward along the given line, but is also changing intrinsically; if initially it was confined to a very small segment of the line, it will, as time goes on, spread itself out over a larger segment: that is to say, we shall be progressively less and less able to predict the result of an observation made to determine the position or the momentum of the particle.

We are now in a position to form some idea of the kind of mathematics that is needed in order to represent quantistic phenomena. The position and momentum of an entity are definite physical quantities, and it is therefore desirable to be able to represent them by single symbols; but these symbols, if they are to express the complexity of the situation and to enable us to make predictions about its future, cannot be ordinary algebraic quantities, any more than symbols representing rotations in three-dimensional space can be ordinary algebraic quantities. We are naturally led, therefore, to look to the new algebras of which Hamilton's calculus was the prototype; just as a single quaternion stands for a combination of four ordinary arithmetic numbers, so we may expect that a symbol representing a physical quantity in quantum theory will represent an aggregate of ordinary numbers.

Now the value of any one of the new algebras depends on our being able to define the operations of addition and multiplication of the symbols in such a way that these operations correspond to ideas or processes of significance in the field of investigation where the algebra is to be applied. Thus we define the multiplication of quaternions in such a way that by its means we can represent and combine rotations in three-dimensional space: in Boolean algebra we define multiplication so as to represent the common part of classes, or the joint assertion of propositions, and we define addition so as to represent the coalition of classes, or the alternative assertion of propositions. Similarly, in the symbolism of quantum theory, sums and products are defined in such a way that in terms of them we can express very simply the computations required for prediction. They are sums and products not of ordinary numbers, but of symbols representing operations. A simple example will show how the idea of an operator may be regarded as an extension of the idea of an ordinary number, and incidentally will provide an illustration of the idea of imperfect definition or blurring.

Let us take two numbers a , b , and multiply each of them by (say) seven, so that we obtain the numbers $a'=7a$, $b'=7b$. This operation performed on two numbers corresponds to the number seven just as much as if it had been performed on a single number. So if we represent the operation by a single symbol, we can fittingly make conventions by which this symbol is equated to seven. If we modify the operation to that represented by the equations $a'=6.99a$, $b'=6.99b$, the symbol can still be equated to a number, namely 6.99; but if we modify it to that represented by the equations

$$\begin{aligned}a' &= 7a + 0.001b, \\ b' &= 7b + 0.001a,\end{aligned}$$

although we can still use a single symbol to represent the operation, it can no longer be equated to a single definite number. The operation, however, differs only slightly from the operation which was represented by seven, and we can regard it as a "blurring" of seven, seeing in it a (not very close) analogy with the blurring which occurs in quantum phenomena. This example suggests that, for the purposes of atomic physics, we should investigate the properties of an algebra whose symbols stand for operations like that last described, in which a set of numbers is transformed into another set of numbers by means of linear equations. Such an algebra was in fact introduced by Cayley not long after the discovery of quaternions; it is called *matrix algebra*, and it has been shown to be capable of representing quantistic phenomena and furnishing predictions regarding them. Non-commutative algebra is, in fact, the symbolism appropriate to things that cannot be measured exactly.

A definite and final answer has thus been given to a question which was asked for many years after the discovery of quaternions, namely

whether a departure from the laws of ordinary algebra was likely to create an instrument of great power for the advancement of physical science. Hamilton himself and his disciples, particularly Tait, solved a vast number of dynamical and physical problems by quaternion analysis, and showed that in many cases the working was remarkably direct intuitive, and concise ; but the problems were classical problems which for the most part had already been solved by classical methods, and there was a general disinclination on the part of mathematical physicists to take the trouble of learning a new calculus until some discovery of the first order had been achieved by its means. The pure mathematicians tended to desert quaternions in favour of the more recently discovered theories of matrices and Boolean algebras, and for many decades the subject languished. Not only so, but heresies sprang up on the physical side. In 1885 Oliver Heaviside in a series of important papers on electromagnetic theory made use of a symbolism which though obviously suggested by quaternions, rejected many of Hamilton's conventions, in particular that which makes the square of a vector negative. The resulting system, which was a kind of hybrid between quaternions and Cartesian mathematics, was not a true calculus, and may perhaps best be described as a syncopated form of classical analysis. When at about the same time another system of the same general type as Heaviside's was proposed by Willard Gibbs in America, a violent controversy broke out between the adherents of the rival algorithms, not unlike that which has raged more recently between those who favour Esperanto, Ido, or Basic English as an international language. The protagonists of quaternions, who were two distinguished Fellows of our Society, P. G. Tait and Cargill Knott, had little difficulty in proving the superiority of the Hamiltonian scheme from the logical and æsthetic points of view ; but the discussion led to no marked improvement in the status of quaternionic studies. It was not until experimental advances had demonstrated the inadequacy of the older physics, and the need for a mathematic capable of grappling with the new situation, that non-commutative algebra came into its kingdom.

The history of Natural Philosophy furnishes some remarkable instances of mathematical theories whose value for the interpretation of the external world was not evident for long ages after their first introduction. Nearly two millennia passed after the discovery of conic sections by Menæchmus before they were identified by Kepler with the orbits of the planets round the sun. In our day things move more quickly, but almost a century elapsed between Hamilton's publication of his ideas in 1843 and the recognition that they lead to a perfect mathematical representation of the physics of the atom.

II. *Note on Tables of an Integral.*

By ALAN FLETCHER, M.A., Ph.D.*

[Received December 9, 1943.]

IN a recent paper in this journal †, Mr. E. O. Powell has given formulæ for computing the integral

$$\int_1^x \frac{\log \xi d\xi}{\xi-1},$$

and has provided a useful table of the integral to 7 decimals, with second differences, for $x=0(0.01)2.00(0.02)6.00$, to use the normal modern notation. From correspondence with the author I understand that the table has been computed without cognizance of two existing but little-known tables, and indeed I am not aware of any tables which duplicate Mr. Powell's values over most of his range. Such comparison with existing tables as is possible may be of interest; it serves to give confidence in the accuracy of Mr. Powell's table.

The works to which I refer are :—

W. Spence, 'An Essay on the Theory of the Various Orders of Logarithmic Transcendents. . .' London and Edinburgh, 1809.

F. W. Newman, 'The Higher Trigonometry. Superrationals of Second Order.' Cambridge, 1892.

Both Spence and Newman develop formulæ of the kind given by Powell. It is, however, with the tables that this note is primarily concerned. The three authors use different notations; we may use Powell's x as above, and refrain from adopting any particular functional notation for the integral.

Newman may then be said to tabulate the integral to 12 decimals for $x=0.50(0.01)1.50$. His values are derived from other quantities, ϕ and ψ , corrected for him by J. C. Adams, and printed on pages 103–104; Newman has merely to form $\phi \pm \psi$. Since comparison with Powell's table revealed errors, Newman's table was checked by re-forming $\phi \pm \psi$ and then differencing. It was found to contain the following nine gross errors, all due to faulty addition or subtraction by Newman and not to errors in the basic values of Adams. The first six corrections refer to page 65 and the last three to page 64 of Newman's book.

When these major errors have been corrected, Newman's table appears to be of good 12-decimal accuracy; the maximum numerical value of

* Communicated by the Author.

† Phil. Mag. xxxiv. p. 600 (1943).

the eighth difference is 100 units of the twelfth place, so that there is no reason to suspect errors of more than a unit or two in that place. Comparison shows that Powell's 7-decimal values in this range are all correct, except that his final digit needs decreasing by unity at $x=1.29$ and increasing by unity at $x=1.42$.

Spence tabulates the integral to 9 decimals for $x=1(1)100$. Thus only five non-zero values, at $x=2(1)6$, can be compared; in addition, the integral is easily evaluated at $x=0$. Comparison suggests that

Powell's x	Newman's x	For	Read	Error
0.55	0.45	5043	5143	1.10^{-2}
0.56	0.44	4411	4361	5.10^{-11}
0.57	0.43	4679	4279	4.10^{-6}
0.59	0.41	0270	0266	4.10^{-8}
0.63	0.37	2133	9133	7.10^{-9}
0.88	0.12	6417	6327	9.10^{-11}
1.37	0.37	0605	7605	7.10^{-9}
1.41	0.41	2861	2857	4.10^{-8}
1.44	0.44	2051	2101	5.10^{-11}

Powell's values are correct, except that his final digit needs increasing by unity at $x=3.00$; I have verified that this increase is required.

Over one hundred of Powell's values, or more than a quarter of the whole, have thus been examined. The three rounding-off errors found all lie between 0.50 and 0.67 of a unit in the seventh decimal place. The claim that "the seventh should not usually be in error by more than one unit" therefore appears to be amply justified.

Articles on tables of the above and related integrals will appear in 'An Index of Mathematical Tables,' by A. Fletcher, J. C. P. Miller and L. Rosenhead (London, Scientific Computing Service, in press).

Mathematical Laboratory,
University of Liverpool.

III. *The Effect of Shape on the Thermal Endurance of Glass Rods*.*

By J. B. MURGATROYD, B.A., F.S.G.T.†

[Received June 21, 1943.]

ABSTRACT.

An examination of published experimental work on the thermal endurance of glass rods makes it possible to deduce a general relationship

* This paper appears also in the Journal of the Society of Glass Technology.

† Communicated by the Author.

between thermal endurance and radius of the form $\theta = \frac{T(1-\mu)}{\alpha E} \cdot \frac{1}{A_0 f(r)}$

in which A_0 is a constant depending only on the glass, and $f(r)$ is a function only of the radius. The expression is derived from the results of tests on borosilicate and on soda-lime-silica glasses, and on others of coefficient of thermal expansion intermediate between these. The experimental results are also used to prepare a table showing the relative maximum stresses developed in rods of different radius subjected to the same thermal shock.

THE INFLUENCE OF ROD RADIUS.

IN the course of a general investigation on the effect of shape on thermal endurance it became necessary to find some expression for the relationship between the thermal endurance of solid cylinders and their radius. The

TABLE I.

Thermal Endurance of Glass Rods. (S.G.T. Committee.)

Radius, cm.	Type of Glass.		
	Borosilicate.	Intermediate.	Soda-lime-silica.
0.1	N.D.*	170°	157°
0.2	293°	130	117
0.3	255	111	101
0.4	N.D.	97	86

TABLE II.

Thermal Endurance of Glass Rods and Linear Thermal Expansion Coefficient (α). (H. Schönborn.)

Radius, cm.	$\alpha=37 \times 10^{-7}$.	$\alpha=55 \times 10^{-7}$.	$\alpha=83 \times 10^{-7}$.
0.20	240°	155°	N.D.*
0.30	195	128	100°
0.40	165	108	97
0.60	130	87	68

results of investigations on the thermal endurance of glass rods published by the Glass Standards Committee of the Society of Glass Technology †, together with those of H. Schönborn ‡, were analysed with this object in view, and the results have shown certain features of general interest.

In Tables I. and II. the thermal endurance values expressed in terms of the mean temperature differences required to break rods of different radius are recorded, and the method of arriving at these values, together

* N.D. means No Determination.

† J. Soc. Glass Tech., Trans. xx. p. 498 (1936).²

‡ *Ibid.* xx. p. 475 (1936).

with a discussion of the tests used in the two sets of experiments, will be found in Appendix I.

When the values given in Tables I. and II. are examined, it is found that a close relationship exists between the thermal endurance of different glasses. This is illustrated in Table III., which gives the ratio of the thermal endurance of one glass to that of another for rods of the same radius.

Thus, in Schönborn's experiments the thermal endurance of the glass of $\alpha=37 \times 10^{-7}$ may be obtained by multiplying the thermal endurance of the glass having $\alpha=83 \times 10^{-7}$ by 1.92. Similarly, the glass of expansion 55×10^{-7} has a thermal endurance which is 1.27 times that of the glass of $\alpha=83 \times 10^{-7}$ at the same radius. The same type of relationship also

TABLE III.

Ratios of Thermal Endurances of Pairs of Glasses of the Same Radius.

S.=Schönborn. S.G.T.=Society of Glass Technology.

Glass Pair.	0.1 cm.	0.2 cm.	0.3 cm.	0.4 cm.	0.6 cm.
S. "37"					
S. "83"	—	—	1.95	1.90	1.91
S. "55"					
S. "83"	—	—	1.28	1.24	1.28
S. "37"					
S. "55"	—	1.55	1.52	1.53	1.61
S.G.T. Borosilicate	—	2.50	2.52	—	—
S.G.T. Soda					
S.G.T. Intermediate	1.08	1.11	1.10	1.12	—
S.G.T. Soda					
S.G.T. Borosilicate	—	2.25	2.30	—	—
S.G.T. Intermediate					

holds for the glasses used by the Glass Standards Committee. For a particular pair over the range of radii tested, the ratio is the same within ± 3 per cent.

Now, this relationship is capable of yielding information of theoretical interest, as the following consideration will show.

Let a thermal shock θ be applied to a rod of radius r made of glass of composition we will call No. 1.

Following Lees *, we can say that the stress set up in the surface of this rod is proportional to the difference between the surface temperature and the mean temperature of the rod.

At some time after the application of the thermal shock this difference will reach a maximum which we may express as $A\theta$, taking care to note

* C. H. Lees, Proc. Roy. Soc. A, ci. p. 411 (1922).

that the fraction A only refers to this particular glass in the form of a rod of radius r , when a thermal shock of value θ is applied. We cannot assume that A will be the same for any other values of r or θ , or for any other glass.

Now let θ be the shock applied which causes the breakage of the rods in question.

Then the stress set up in the surface * is

$$\frac{A\theta\alpha E}{1-\mu},$$

where α =linear coefficient of expansion.

E=Young's modulus.

μ =Poisson's ratio.

Since the rods break, this stress must equal the tensile strength, T, of the glass †, i. e.,

$$\frac{A\theta\alpha E}{1-\mu} = T$$

or

$$A\theta = \frac{T(1-\mu)}{\alpha E}.$$

The right-hand side of this equation consists entirely of constants representing physical properties of glass A. Let us, for convenience, call it ϕ .

Then $A\theta = \phi$ (1)

Similarly, for rods of a glass of composition No. 2, and the same radius, r , which break at an applied thermal shock of θ' ,

$$B\theta' = \phi',$$

where B is the particular fraction of θ' defined as closely as the fraction A, and ϕ' is $\frac{T(1-\mu)}{\alpha E}$ for the glass B.

Let us also consider glass No. 1 and glass No. 2 tested in the form of rods of a different radius r_1 . The temperature difference required to break rods of glass composition No. 1 is now θ_1 , and the maximum difference between the mean rod temperature and the surface temperature can be written

$$A_1\theta_1,$$

where A_1 is the particular fraction for glass No. 1 corresponding to the new temperature θ_1 and new radius r_1 .

* Ignoring for the moment the possibility of elastic deformation.

† See V. H. Stott, J. Soc. Glass Tech., Trans. viii. p. 139 (1924); and C. E. Gould and W. M. Hampton, *ibid.*, Trans. xiv. p. 188 (1930).

But as the rods break *, then, as pointed out previously,

$$A_1\theta_1=\phi.$$

Similarly, for glass No. 2 in the form of rods of radius r_1

$$B_1\theta_1'=\phi',$$

where θ_1' is the breaking temperature for these rods. If we consider these expressions in relation to the experimental results, we note that at all values of the radius the ratio of the thermal endurance of glass No. 1 to glass No. 2 is constant, i. e.,

$$\frac{\theta}{\theta'} = \frac{\theta_1}{\theta_1'} = \text{constant.}$$

But we have seen that

$$\theta = \frac{\phi}{A}, \quad \theta_1 = \frac{\phi}{A_1}, \quad \theta' = \frac{\phi'}{B}, \quad \theta_1' = \frac{\phi'}{B_1}.$$

Hence

$$\frac{\frac{\phi}{A}}{\frac{\phi'}{B}} = \frac{\frac{\phi}{A_1}}{\frac{\phi'}{B_1}} = \text{constant,}$$

or

$$\frac{\phi}{\phi'} \cdot \frac{B}{A} = \frac{\phi}{\phi'} \cdot \frac{B_1}{A_1} \dots = \text{constant.} \quad (2)$$

In other words, to comply with the experimental results, the ratio $\frac{B}{A}$ must be the same at any radius.

Hence, the law relating the fraction A to the radius is the same as that which governs the fraction B, and is such that if A_0 and B_0 are the fractions for some particular radius, then at radius r , $A=f(r)A_0$, $B=f(r)B_0$, $C=f(r)C_0 \dots$, where $f(r)$ represents a function of the radius only. Hence, part of the fraction A as originally defined can be regarded as a constant for all radii, and related only to the glass which is being examined. We can therefore express the thermal endurance of the glass we have called No. 1 as follows :

$$f(r)A_0\theta=\phi,$$

or

$$f(r)\theta = \frac{\phi}{A_0}.$$

Similarly, for glass No. 2,

$$f(r)B_0\theta'=\phi',$$

or

$$f(r)\theta' = \frac{\phi'}{B_0},$$

and $f(r)\theta'' = \frac{\phi''}{C_0}$ for a third glass, and so on ; and as ϕ , ϕ' , etc., A_0 , B_0 ,

* It is assumed that, within the range of radii covered by thermal endurance tests, there is no change in the tensile strength of these glasses.

etc., are constants, we can write the general expression for the thermal endurance of glass rods as

$$\theta = \frac{Q}{f(r)}, \quad \dots \dots \dots (3)$$

where θ is the average breaking temperature of any type of glass in the form of rods, and Q is a constant relating only to the glass, $= \frac{T(1-\mu)}{\alpha E} \cdot \frac{1}{A_0}$ and $f(r)$ is a function relating only to the radius.

Now, at this stage two assumptions are to be made. The first is that A_0 remains constant beyond the range of radii used in the experimental work; the other is that $f(r)$ approaches unity as r becomes very large. In justification of the first assumption it can be said that the range tested covers a variation of 100 per cent. in the thermal endurance values of a glass, and in this respect, therefore, is reasonably wide; also, the changes in thermal endurance as the radius is increased beyond the range tested are already becoming small in Schönborn's experiments at least.

The second assumption is merely another way of saying that curvature has no effect on thermal endurance when the radius is very large. Such an assumption is in accordance with our knowledge of the subject.

Keeping the assumptions in mind, therefore, we can test expression (3) against the experimental results. The value of $\frac{T(1-\mu)}{\alpha E} \cdot \frac{1}{A_0}$ which, for convenience, has been written as Q , now becomes the thermal endurance of very large rods. Hence, if we can determine Q experimentally for all the glasses, we should find that the ratio $\frac{\theta}{Q}$ should be the same for them all at the same radius, *i. e.*, should be $\frac{1}{K_r}$ (where K_r is the particular value of $f(r)$ at that radius).

The S.G.T. tests were not carried out on sufficiently large rods to enable Q to be determined by extrapolation, but Schönborn tested rods sufficiently thick to indicate the limiting thermal endurance of very large rods. The values obtained in this way for Schönborn's glasses were:

Glass.	Thermal Endurance
	as $r \rightarrow \infty$.
$a=37 \times 10^{-7}$	115°
$a=55 \times 10^{-7}$	75
$a=83 \times 10^{-7}$	60

The values of $\frac{\theta}{Q}$ for each of these glasses at different radii, as obtained from the experimental results, are shown in Table IV.

In order to obtain some check on the S.G.T. figures, cylinders 5 cm. in diameter and 5 cm. in length were prepared from a soda-lime-silica

glass similar to that used in the S.G.T. tests. The approximate thermal endurance of these was found to be 60° , a value which was therefore taken to represent Q for the S.G.T. glass. Values of $\frac{\theta}{Q}$ obtained from the experiments are shown in Table V.

The agreement between Schönborn's results and the S.G.T. results is good at 0.3 cm. and 0.4 cm. radius, but there is a divergence at 0.2 cm. radius. This could be due to a change in A_0 or $f(r)$ at this radius consequent on the different methods of test, or it might represent the greater chance of experimental error which occurs in measuring thermal endurance

TABLE IV.

Values of $\frac{\theta}{Q}$ at Different Radii. (Calcd. from Schönborn.)

Radius, cm.	$\frac{\theta}{Q}$		
	Glass. $a=37 \times 10^{-7}$.	Glass. $a=55 \times 10^{-7}$.	Glass. $a=83 \times 10^{-7}$.
0.2	2.09	2.07	—
0.3	1.70	1.71	1.66
0.4	1.43	1.44	1.45
0.6	1.13	1.16	1.13

TABLE V.

Values of $\frac{\theta}{Q}$ for S.G.T. Soda-Lime-Silica Glass.

Radius, cm.	$\frac{\theta}{Q}$
0.2	1.95
0.3	1.70
0.4	1.43

when it reaches a relatively high value. There seems to be some justification for further experimental work to determine if the general expression $\theta = \frac{Q}{f(r)}$ does, in fact, represent the relationship between the thermal endurance of glass rods and their radius. The values of $\frac{\theta}{Q}$ are shown plotted against radius in fig. 1. This represents the curve of $\frac{1}{f(r)}$ plotted against r (since $\frac{\theta}{Q} = \frac{1}{f(r)}$).

CALCULATION OF THE STRESSES DEVELOPED BY THERMAL SHOCK.

From fig. 1 it is possible to construct a table showing the relative stress developed when rods of different radii are subjected to thermal

shock. The value of Q represents the thermal shock required to break rods of very large radius. For rods of 0.6 cm. radius the thermal shock required to cause fracture is $1.14 \times Q$, and, for radius 0.4 cm., $1.44 \times Q$, and so on. Hence, the stress developed by a shock Q when the radius approaches infinity can be written as 1, and the stress developed by

Fig. 1.

Relation between $\frac{\theta}{Q}$ for rods of different radius.

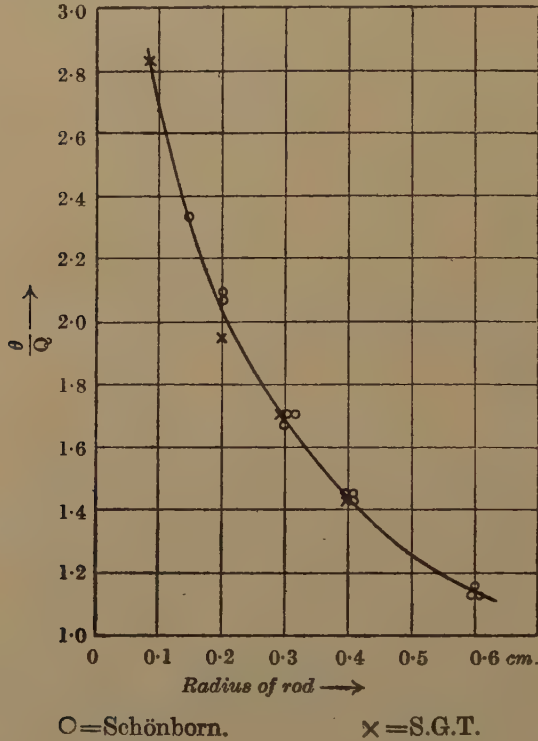


TABLE VI.

Relative Stress Developed by the Same Thermal Shock Applied to Rods of Different Radius.

Radius, cm.	Relative stress.	Radius, cm.	Relative stress.
∞	1	0.4	0.70
0.6	0.88	0.2	0.48

shock Q at radius = 0.6 cm. is $\frac{1}{1.14} = 0.88$, and similarly for other radii.

Table VI. shows the relative stresses developed by a thermal shock on rods of different radius, derived in this manner.

The equation relating $\frac{\theta}{Q}$ and radius is not a simple one, and the writer prefers to present the curve of values of $\frac{\theta}{Q}$ at different radii, as in fig. 1, until further investigation of the relationship has been undertaken, rather than attempt to express $f(r)$ as an equation.

THE NATURE OF THE CONSTANTS, A_0 , B_0 , ETC.

In conclusion, it is of interest to examine the full implications of the constant which has been described as A_0 , B_0 , . . . , etc. If it can be shown that this factor is truly constant for a particular glass—and the analysis of the two sets of results already made suggests that it may well be so—then certain other interesting conclusions can be drawn. It will be remembered that A , B , . . . were defined as the fractions of the applied thermal shock, θ , which became directly responsible for producing stress in the surface of the rods. In the definition the value of A was restricted to that which operated at one particular radius, and at one particular value of θ ; but the experimental results showed that A changed only when the radius was changed, and that even then it changed in the same ratio for all glasses. Now, this factor A includes in itself all the factors which affect the stress system set up when a thermal shock is applied to a rod. These are: the radius of the rod, the diffusivity of the glass, the rate of heat transfer from glass to chilling medium, and the stress multiplication due to the "end effect" in the rod. All these must be taken into consideration when we attempt to calculate the stress set up by a thermal shock. So far we appear to have taken no account of them, *but actually we have implied their effect in the factor A* . The latter is, therefore, not so simple as it appeared to be, and the final assessment of the interplay between the variation of chilling efficiency, the "end effect," and the temperature distribution due to curvature and diffusivity, may all be expressed in the function $f(r)$. We are then faced with two somewhat unexpected results. (1) The complicated interaction of all the factors described is, in their total effect, related only to the radius; and the relation is the same for the different glasses examined so far. (2) The diffusivity of the glasses apparently has little effect on the maximum stress which occurs in the surface of a glass rod when a thermal shock is applied, although, naturally, it may affect the time taken to reach the maximum stress. This deduction may be derived from the fact that the diffusivities of the soda-lime-silica and the borosilicate glasses differ by about 25 per cent., yet the values of $\frac{\theta}{Q}$ are apparently the same at the same radius.

In the writer's opinion, the possibility that the thermal endurance of glass rods can be expressed generally by a formula of the type $\theta = \frac{Q}{f(r)}$ is worth further investigation, as it would be useful when comparisons

are to be made between the thermal endurances of different glasses, if found to be generally true. The theoretical significance of a "universal" function related only to radius has already been discussed, and the unexpected nature of the result pointed out. The reader is referred to

TABLE VII.

Experimental Determinations of Thermal Endurance of Glass Rods by

Different Investigators Obey the Law $\frac{\theta}{Q} = \frac{1}{f(r)}$.

Values of $f(r)$ used in this table were taken from fig. 1.

S.G.T. Soda-Lime-Silica Glass.				S.G.T. Intermediate Glass.			
Radius, cm.	$f(r)$.	θ deter- mined.	$\theta f(r)$.	Radius, cm.	$f(r)$.	θ deter- mined.	$\theta f(r)$.
0.1	0.37	157°	58°	0.1	0.37	170°	63°
0.2	0.49	117	57	0.2	0.49	130	64
0.3	0.59	101	59	0.3	0.59	111	65
0.4	0.70	86	60	0.4	0.70	97	68

S.G.T. Borosilicate Glass.				Schönborn $\alpha=37 \times 10^{-7}$.			
Radius, cm.	$f(r)$.	θ deter- mined.	$\theta f(r)$.	Radius, cm.	$f(r)$.	θ deter- mined.	$\theta f(r)$.
0.2	0.49	293°	144°	0.2	0.49	240°	118°
0.3	0.59	255	150	0.3	0.59	195	115
				0.4	0.70	165	116
				0.6	0.88	130	114

Schönborn $\alpha=55 \times 10^{-7}$.				Schönborn $\alpha=83 \times 10^{-7}$.			
Radius, cm.	$f(r)$.	θ deter- mined.	$\theta f(r)$.	Radius, cm.	$f(r)$.	θ deter- mined.	$\theta f(r)$.
0.2	0.49	155°	76°	0.3	0.59	100°	59°
0.3	0.59	128	76	0.4	0.70	87	61
0.4	0.70	108	76	0.6	0.88	68	60
0.6	0.88	87	76				

Tabata and Moriya's Borosilicate Glass.

Radius, cm.	$f(r)$.	θ determined.	$\theta f(r)$.
0.15	0.43	238°	102°
0.20	0.49	210	103
0.25	0.54	186	100
0.30	0.59	173	102

W. M. Hampton's general discussion of the thermal endurance of glass * for a better appreciation of this. Hampton and Gould derived a general thermal endurance equation of the form

$$\theta = \frac{T(1-\mu)}{\alpha E} \cdot \frac{1}{f(xt)},$$

* J. Soc. Glass Tech., Trans. xx. p. 461 (1936).

where $\frac{1}{f(xt)}$ is a function involving size, shape, diffusivity, and rate of heat transfer from glass to chilling medium. In this paper experimental values for the thermal endurance of glass rods have been examined and the conclusion reached that, for this shape, $\frac{1}{f(xt)}$ may possibly be reduced to a function related only to radius, *i. e.*, $\frac{1}{A_0 f(r)}$, where A_0 is a constant for a particular glass.

In what manner the combined effects of curvature, diffusivity, rate of heat transfer to chilling medium, and stress-increase due to end-effect, could become reduced to a simple function of this nature is discussed in Appendix II. As a final indication that a close approximation to such an effect occurs in practice Table VII. is presented. This gives the values of Q determined for each glass mentioned in Table I. and, in addition, for a borosilicate glass investigated by K. Tabata and T. Moriya * by a method similar to that used by the Glass Standards Committee. The

value of $f(r)$ at each radius quoted was obtained from fig. 1 (in which $\frac{\theta}{Q} = \frac{1}{f(r)}$); the experimentally determined thermal endurences were then multiplied by $f(r)$, the result giving a constant for each glass, *i. e.*, the value of Q for that glass. It can be seen from Table VII. that the thermal endurance of all the glasses can be expressed by a formula of the type $\theta = \frac{1}{f(r)}$ with a reasonable degree of accuracy.

APPENDIX I.

Discussion of the Thermal Endurance Tests Employed.

In the tests described by the Glass Standards Committee the rods used were 1 inch in length, had the ends fire-polished, and, after the fire-polishing, the rods were annealed. In the test the rods were maintained at the desired temperature in an electrically heated furnace, and were then removed by a "rake" and dropped immediately into a bath of cold water. This process was repeated for each rod until breakage occurred, and the mean breaking temperature was then recorded for rods of the same diameter. The "steps" by which the thermal shock was increased were 5°.

The values of θ were presented in the report in the form of graphs from which Table I. in this paper was compiled. An exception to this procedure was made in the case of the borosilicate glass, which gave widely varying results. The two values given for this glass were obtained from the recommendations of Dr. Seddon after he had subjected the

* J. Amer. Ceram. Soc. xiii. p. 34 (1934).

results to statistical analysis. The reader is referred to Dr. Seddon's report for the method by which this selection was made.

For the soda-lime-silica glasses experiments were extended by the author to include rods of radius down to 0.085 cm., the rods being of the same manufacture and composition as those used by the Glass Standards Committee. Using the same methods the thermal endurance of rods of this radius was found to be 172°, and this value was used to extend the graph presented in the report.

Schönborn's experiments were made on fire-polished and annealed glass rods, 12.0 cm. in length, which were held in an electrically heated furnace to attain the desired temperature. They were then removed rapidly by insulated tongs and dipped into cold water covering three-quarters of their length, and rapidly moved about to assist in the transfer of heat from the glass to the water. The "steps" by which the thermal shock was increased were 10°, but the breaking temperature was recorded as the average value between the temperature at which fracture occurred and the last previous temperature successfully withstood.

In principle the two tests are similar: in each the rod is heated in air and quenched in water, and breakage of technically perfect rods will commence at the end, where the stresses are increased by the readjustment of the bending moments. The chief difference between the two methods lies in the fact that Schönborn assisted heat transfer by agitating the water around the rods; the heat transfer should therefore have been more efficient, and the thermal endurances recorded should be somewhat lower than if the S.G.T. method had been used.

Schönborn presented his results in the form of graphs from which the values recorded in Table I. were obtained. It should be noted that these values are recorded as "technical" results by Schönborn. This means that they represent the actual breaking temperatures without any selection, and correspond therefore with the Glass Standards Committee results. Schönborn regarded some of his test breakages as due to imperfections in the glass, and preferred to ignore those rods which fractured along the length of the rod and through an end. The author is not in agreement with this method of selection of results, and indeed would regard the type of fracture described as being normal for a glass rod. The "end-effect" is likely to increase the tangential stress at the end, thereby causing fracture in a radial direction through the end. For this reason alone, therefore, Schönborn's "technical" results would be considered more representative of the true thermal endurance of the glass; in addition to this no such selection is made by other investigators, and it should not be done therefore with Schönborn's glasses when comparisons are to be made.

It may be noted that Schönborn's graphs as reproduced are somewhat small, so that it is only possible to determine the thermal endurances from them with an accuracy of about ± 3 per cent., although this order of accuracy is sufficient for the work of the type described in this paper.

APPENDIX II.

The Factors Affecting Thermal Endurance.

The deductions to be made from the fact that rods of glasses of different composition have thermal endurences which obey the law $\theta = \frac{Q}{f(r)}$ are :

1. The diffusivity has no effect on the shape of the temperature distribution curves which occur from instant to instant immediately after the hot rods are plunged into the chilling medium ; it merely affects the time at which a particular temperature distribution occurs. This is not immediately obvious from a consideration of the conditions ; for the central axis of the rod is not maintained at a constant temperature by a heat source. So soon as the "chilling wave" reaches the centre the temperature there begins to drop. Hence one would expect that, if the rate of chilling of the surface were not very rapid compared with the rate of cooling within the rod, a condition might arise wherein the temperature of the centre of the rod began to fall before the outside surface had been chilled to 0° . In such circumstances the temperature distribution would favour resistance to fracture, and a glass with a sufficiently high value of diffusivity would show abnormally good thermal endurance.

It must be concluded, therefore, that the composition of the glass does not affect appreciably the rate at which the surface chills in water ; and that none of the glasses tested so far have sufficiently large diffusivities to introduce the effect discussed above.

2. Having admitted the conclusions drawn in (1) above, the temperatures of rods of a particular radius pass through the same series of distributions from the instant of immersion in all the glasses tested. If this is the case the stress-multiplication due to the "end-effect" will be the same in all the rods ; for the shape of the temperature distribution curve is the deciding factor as to how much bending is permissible.

From the above one can conclude that all the glasses tested obeyed the same surface cooling law, and that none of their diffusivities was large enough to upset the shape effect. The "end-effect" produced the same percentage stress-increase because the temperature distributions at the instant of maximum stress were the same in all the glasses. Hence the only variable left was shape, *i. e.* the radius of the rods.

The conclusion then follows that the value of A_0^* should be the same for all the glasses, because it now merely represents the effect due to the rate of cooling of the surface and the stress increase due to bending ; and we have seen that this must have been the same for all the glasses. Unfortunately, this cannot be checked for the particular glasses referred to in the text, because the necessary values of T , μ , α , and E , are not known.

$$* Q = \frac{T(1-\mu)}{\alpha E} \cdot \frac{1}{A_0}.$$

IV. Thermal Diffusion in Mixtures of Molecules of Small Mass Difference.

By K. E. GREW, Ph.D.*

[Received October 12, 1943.]

SUMMARY.

The thermal diffusion effect has been examined in gas mixtures in which the difference in the molecular masses is small and the heavier molecule has the smaller diameter. In such cases Chapman (1940) has shown from theoretical considerations that the thermal diffusion factor α may be negative (the heavier molecule then diffuses up the temperature gradient), or it may change sign with changing proportions of the constituents. This change of sign was found to occur in Ne-NH₃ mixtures; it did not occur in A-HCl mixtures. N₂O-CO₂ mixtures were also examined, since the experimental data of viscosity and diffusion suggest that here too a reversal of sign of α may occur with changing composition. No reversal was found.

Introduction.

THE general expression for the thermal diffusion ratio k_T in its first approximation is (Chapman and Cowling, 1939)

$$[k_T]_1 = 5(C-1) \frac{S_1 n_{10} - S_2 n_{20}}{Q_1 n_{12} + Q_2 n_{21} + Q_{12}}, \quad \dots \quad (1)$$

where n_1, n_2 = number density of the molecules of kinds 1 and 2;

$$n_0 = n_1 + n_2; \quad n_{12} = n_1/n_2; \quad n_{21} = n_2/n_1; \\ n_{10} = n_1/n_0; \quad n_{20} = n_2/n_0.$$

$C, S_1, S_2, Q_1, Q_2, Q_{12}$ are quantities depending on the force law and the molecular masses.

Chapman (1940) has discussed in detail the implications of this expression, which he re-writes as

$$[k_T]_1 = n_{10} n_{20} [\alpha]_1 = n_{10} n_{20} 5(C-1) \frac{S_1 n_{10} - S_2 n_{20}}{Q_1 n_{10}^2 + Q_2 n_{20}^2 + Q_{12} n_{10} n_{20}}, \quad \dots \quad (2)$$

where $[\alpha]_1$ is the first approximation to the thermal diffusion factor α . The terms involving Q are positive probably for all physically suitable molecular models, and certainly when the force field is of the inverse power type, $F = \kappa r^{-\nu}$. In this case the factor $(C-1)$ also is positive if the mutual force index ν_{12} is greater than 5, as is usually the case. The

* Communicated by the Author.

sign of α thus depends on S_1 and S_2 . The signs of S_1 and S_2 in turn depend in a complex manner on the mass difference, the force index and the molecular diameters. Generally it may be said that for mixtures in which the mass difference ($m_1 - m_2$) is large and positive, S_1 is positive and S_2 negative, so that α is positive at all concentrations. This means that the heavier molecule diffuses down the temperature gradient. This is the case commonly observed. When, however, ($m_1 - m_2$) is small (but still positive) the molecular diameters play a determining rôle, and interesting possibilities arise. Firstly, if, contrary to the usual condition, the heavier molecule is the smaller, S_1 may be negative and S_2 positive, so that α has a negative value. In this case the heavier molecule should diffuse up the temperature gradient. Secondly, S_1 and S_2 may be of the same sign; then α should change sign, and therefore vanish, at some particular concentration.

Mixtures have been examined in which one or both of these effects might be expected to occur. They are A-HCl and Ne-NH₃. In the first the molecular mass ratio $m_1/m_2 = 1.095$ and (viscosity) diameter ratio $s_1/s_2 = 0.815$. In the second $m_1/m_2 = 1.186$ and $s_1/s_2 = 0.587$. If the molecules are treated as rigid elastic spheres, the calculated value of α is positive at all concentrations for A-HCl, and negative at all concentrations for Ne-NH₃. N₂O-CO₂ mixtures were also examined because, as Chapman (1940, p. 59) has shown in discussing the case in which ($m_1 - m_2$) = 0, the experimental data for these gases suggest that here too α should show a reversal of sign as the composition changes.

The values of α in all these mixtures are very small. Thus for A-HCl, when $n_{10} = n_{20} = 0.5$, $[\alpha]_1 = 0.0036$, and therefore $k_T = 0.0009$. The actual values are likely to be about half these for rigid elastic spheres. Direct measurement of the separation due to thermal diffusion was therefore not attempted, but the mixtures were treated in a Clusius-Dickel column, in which the separation due to thermal diffusion is multiplied by convection. Although absolute values of α are not readily found from the measurements made in this way, the sign of α is indicated by the direction of the longitudinal concentration gradient set up in the column, since this is determined by the direction of the transverse thermal diffusion.

Method.

The Separating Column.—This consisted of two coaxial pyrex glass tubes enclosed in a brass tube forming a water-jacket. The inner tube, which was sealed to the outer glass tube at the upper end, was filled with aniline which could be boiled by means of an electric heater immersed in it: a uniform temperature over the whole length of the tube was thus obtained. The outer tube was cooled by a flow of water at room temperature. The dimensions were

inner tube: 18.2 mm. external diameter; length 140.5 cm.

outer tube: 30 mm. internal diameter; length 144.5 cm.

Preparation of the Gases.—The neon and argon used were supplied by the British Oxygen Co. as spectroscopically pure. Ammonia was prepared from lime and ammonium chloride. It was dried over ignited lime, condensed, and the middle fraction collected. HCl was prepared by dropping hydrochloric acid into pure concentrated sulphuric acid, drying the gas over phosphorus pentoxide, condensing and fractionating. Nitrous oxide and carbon dioxide were taken from cylinders as supplied for anæsthetic purposes; both were dried, then condensed by means of a liquid oxygen bath, and the middle fraction collected.

Analysis.—The composition of the gas at both ends of the tube was determined chemically. Samples of about 4 cm.³ (measured at atmos. pressure) were withdrawn alternately from the upper and lower ends of the column. They were admitted to a burette, where the pressure before and after absorption of one of the constituents could be determined. NH₃ was absorbed by a drop of H₂SO₄ on a sintered glass bead; HCl by a pellet of KOH; CO₂ in the same way. The error in the analysis was less than 0.5 per cent. of the percentage composition.

The gas mixtures were initially at a pressure of about 54 cm. of mercury, but the pressure fell as samples were withdrawn for analysis. The average pressure during the separation process was roughly 45 cm. There was no indication that the separation was affected by the pressure change. Analyses were made only after the heater had been on for two hours or more, though equilibrium was probably attained much sooner; Wall and Holley (1940), using a longer column, found 20 minutes sufficient. The inner tube was at the temperature of boiling aniline, 184° C., while the mean temperature of the outer varied from 13 to 17° C. The mean temperature of the gas as determined from the pressure change when the heater was set in operation was 64° C. when the water-jacket was at 17° C.

Results.

The thermal separation, *i. e.* the difference in percentage composition by volume at the bottom and top of the tube is shown as a function of the composition in fig. 1. In A-HCl mixtures the heavier molecule concentrates at the lower end of the column; it therefore diffuses down the temperature gradient and the thermal diffusion factor α is positive at all compositions. The Ne-NH₃ mixtures, however, show a change in the direction of the concentration gradient. From 0 to 75 per cent. Ne the Ne molecules concentrate at the top of the column, which means that they are diffusing up the temperature gradient and α is negative over this range, while from 75 to 100 per cent. Ne the reverse holds. In N₂O-CO₂ mixtures the N₂O concentrates at the lower end at all compositions.

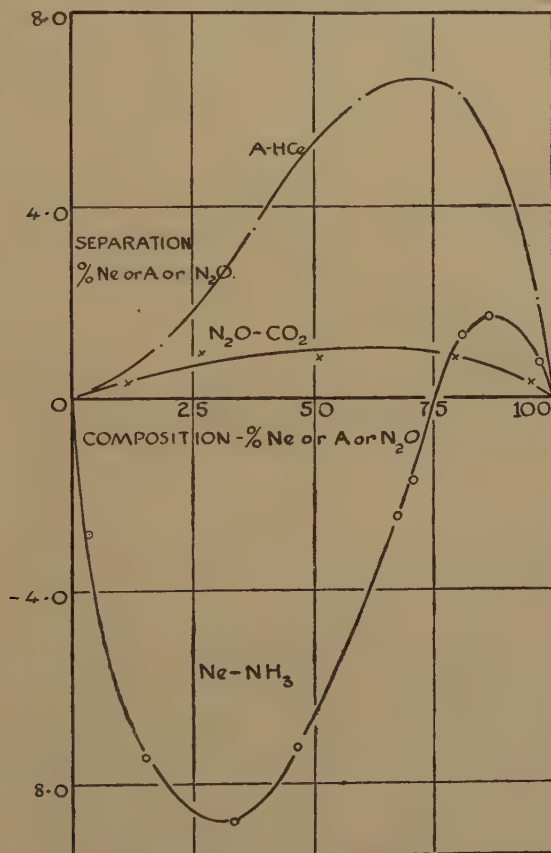
Discussion.

The results show that Ne-NH₃ is one of what must be few mixtures in which the diameter difference is a factor dominating the mass difference

in the thermal diffusion process. Moreover, the two factors are so related that not only are negative values of the thermal diffusion factor observed *, but the mixture exhibits a change of sign of α with varying composition. Its behaviour therefore illustrates qualitatively the possibilities discussed by Chapman (1940).

A satisfactory quantitative comparison of the experimental results with the theoretical expressions is not possible, because, firstly, the

Fig. 1.



The separation (*i. e.* concentration at the bottom of column minus that at the top) as a function of the composition.

measurements do not yield α directly, and, secondly, the theoretical expression for $[\alpha]_1$ involves not only the viscosity diameters s_1 and s_2 but also the "mutual" diameter s_{12} . In the case of Ne-NH₃, however, an attempt can be made as follows. The quantities S and Q in the expression (2) for $[\alpha]_1$ are functions of s_1 , s_2 , and s_{12} ; s_1 and s_2 may be

* Leaf and Wall (1942) have found α to be negative in CO₂-cyclopropane mixtures.

taken as known from viscosity data; s_{12} can then be determined if use is made of the fact that there is no separation in a mixture containing 75 per cent. Ne. At this composition then α is zero and therefore $3S_1=S_2$. From this relation s_{12} can be determined in terms of a quantity A , which depends only on the mutual force law (Chapman, 1940, p. 40). Assuming now that the molecular interaction is an inverse power repulsion, $S_1, S_2, Q_1, Q_2, Q_{12}$ can be calculated for various given values of the mutual force index ν_{12} . This was done for values of ν_{12} equal to 7, 9 and 15. The determination of values of α from the separation measurements, for comparison with these calculated values, involves the theory of the column. This has been given by Furry, Jones and Onsager (1939) on the assumption that the gas can be regarded as confined between plane parallel surfaces. The expression they find is, ignoring a factor which is nearly unity,

$$\alpha = \frac{4A_1 \rho g w^4}{63 \eta D_{12}}$$

in which ρ = mean density of the mixture at the mean temperature.

η = „ viscosity „ „ „ „

D_{12} = coefficient of diffusion „ „ „

and A_1 is given by

$$e^{4A_1 l} = \frac{n_{20}^U n_{10}^L}{n_{10}^U n_{20}^L} = \text{separation factor } F.$$

Here $n_{10}^{U(L)}$ = concentration of molecules 1 at upper (lower) end.

$n_{20}^{U(L)}$ = „ „ 2 „ „

$2l$ = length of column.

$2w$ = width of annulus containing the gas.

Wall and Holley (1940) have used this expression to calculate α for various mixtures, including N_2O-CO_2 . For the present work, D_{12} being unknown, we have calculated only $\log F/\eta$, to which α should be proportional if the dependence of D_{12} and ρ on the concentration is disregarded. Approximate values of η were found from the expression given by Chapman and Cowling, p. 230. The curve of $\log F/\eta$ (after multiplication by a constant to make the value for a 50 per cent. mixture coincide with that of $[\alpha]_1$) as a function of the composition is shown in fig. 2, together with that of the calculated value of $[\alpha]_1$ for $\nu_{12}=9$. The curves are very different. Those for $\nu_{12}=7$ and $\nu_{12}=15$ give no better agreement.

If the theory of the column is valid, the dissimilarity of the theoretical and experimental curves is probably an indication that the inverse power force law, on which the calculation of $[\alpha]_1$ was based, is inadequate to represent the molecular interactions. Apart from the evidence of viscosity data for a long range attractive field in ammonia, there is now the recent work of Watson and Woernley (1943), who have examined thermal diffusion in $N^{14}H_3-N^{15}H_3$ mixtures. They find a rapid decrease of α as the temperature is lowered, with a change from positive

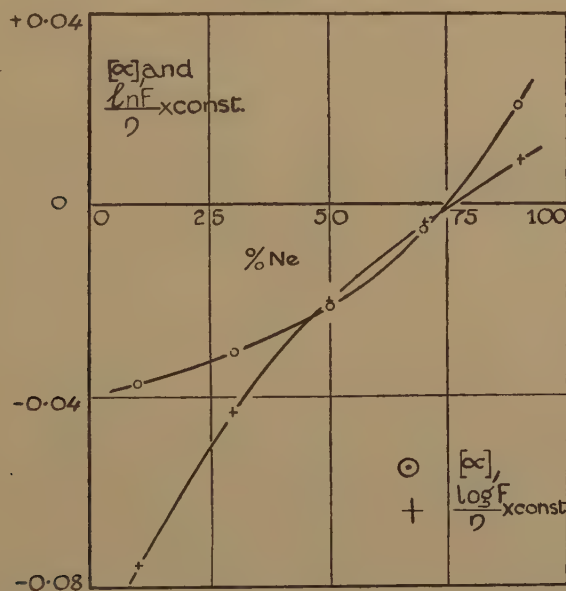
to negative values at about room temperature. Clark Jones (1941) has shown that such a variation is to be expected for molecules of the Sutherland or Lennard-Jones types.

The $\text{N}_2\text{O}-\text{CO}_2$ mixture has been discussed by Chapman (1940, p. 59) as an example of the special case when the mass difference (m_1-m_2) is zero. In such a case α changes sign with varying composition if two quantities

$$x_{12} = \frac{5A}{2} \frac{s_{12}^2}{s_1^2}, \quad x_{21} = \frac{5A}{2} \frac{s_{12}^2}{s_2^2}$$

lie on the same side of unity. The experimental data for this mixture, given in his Table 2, p. 45, indicate that, if the molecular interaction is

Fig. 2



Ne- NH_3 : variation with composition of $[a]_1$ calculated for $\nu_{12}=9$, and of $\log F/\eta \times \text{constant}$.

of the inverse power type, this condition is here fulfilled. Since no reversal of sign occurs, x_{12} and x_{21} must lie on opposite sides of unity. If the diameters s_1 , s_2 and s_{12} are correct, A must therefore have a value of about 0.47. The corresponding value of ν_{12} , assuming an inverse power force law, is about 10. This is improbably high, and it seems that in this case also an inverse power repulsive force is not adequate to represent the molecular interaction.

I wish to thank Professors S. Chapman and R. Fürth for their suggestions and criticism in connection with this work.

References.

- *Chapman, S., Proc. Roy. Soc. A, clxxvii. p. 38 (1940).
 Chapman, S., and Cowling, T. G., 'The Mathematical Theory of Non-uniform Gases.' Camb. Univ. Press (1939).
 Furry, W. H., Jones R. Clark, and Onsager, L., Phys. Rev. lv. p. 1083 (1939).
 Jones, R. Clark, Phys. Rev. lix. p. 1019 (1941).
 Leaf, B., and Wall, F. T., J. Phys. Chem. xlv. p. 820 (1942).
 Wall, F. T., and Holley, C. E., J. Chem. Phys. viii. p. 949 (1940).
 Watson, W. W., and Woernley, D., Phys. Rev. lxiii. p. 181 (1943).

Heriot-Watt College,
 Edinburgh.

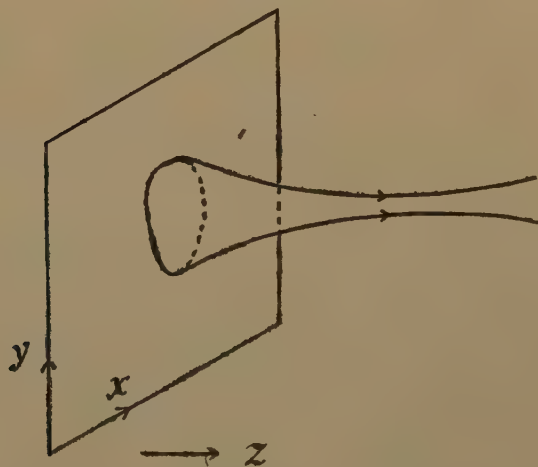
V. Effect of Space Charge on Electron Beams.

By J. AHARONI, Department of Physics, King's College, London †.

[Received January 15, 1943.]

THE effect of the space charge on an electron beam as it emerges from an electron-optical system has been treated in a number of papers (see bibliography). In the following a new method of dealing with the

Fig. 1.



subject is described and a few applications are carried out. The method may be regarded as a hydrodynamical method. As we are dealing with stationary beams the time can be eliminated in a convenient way from the calculation.

* Professor Chapman points out the following misprints in this paper: in (4.1) for $M_1 M_2$ read $(M_1 M_2)^\dagger$; in (4.6) delete $M_1 M_2$; in (10.2) for $c_1 c_2 Q_2$ read $c_1 c_2 Q_{12}$. Also in Table 2 the value of x for $N_2 O - CO_2$ should be 1.11, not 1.05.

† Communicated by Dr. H. T. Flint.

Consider a beam of electrons emerging from a hole in a plane $z=0$ with a given space charge density $\rho_0(x, y)$ in the plane of the hole and a given velocity distribution $U_0(x, y)$. If the electrons did not act upon one another, each element of charge would continue to flow in the direction of its initial velocity, and the beam would converge or diverge as the case may be, in the former case the beam shrinking to an infinitesimal volume, where the charge density would become infinite. Actually this will in general not happen as the electrons repel one another. This is what is termed the space charge effect.

Maxwell's equations are :—

$$\text{curl } \mathbf{E} = -\dot{\mathbf{H}},$$

$$\text{curl } \mathbf{H} = \mathbf{J} = \rho \mathbf{U} + \dot{\mathbf{E}},$$

$$\text{div } \mathbf{E} = \rho,$$

$$\text{div } \mathbf{H} = 0,$$

where ρ is the density of the charge in a vacuum, \mathbf{U} the velocity of the charge, \mathbf{J} =total current density. In addition,

$$\mathbf{F} = \rho(\mathbf{E} + [\mathbf{U}, \mathbf{H}]) = \text{force per unit volume.}$$

Since we are considering stationary beams

$$\dot{\mathbf{E}} = \dot{\mathbf{H}} = 0.$$

The charge ρ per unit of volume will have a mass μ per unit of volume and

$$\frac{\mu}{\rho} = \frac{e}{m} = \text{the specific charge.}$$

The dynamical eqn. of the moving charge will be

$$\mu \frac{D\mathbf{U}}{Dt} = \rho(\mathbf{E} + [\mathbf{U}, \mathbf{H}]). \quad \dots \dots \dots (1)$$

$\frac{D}{Dt}$ means the rate of change of a quantity while moving with the charge mass element.

For a scalar α

$$\frac{D\alpha}{Dt} = \frac{\partial \alpha}{\partial t} + \frac{\partial \alpha}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial \alpha}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial \alpha}{\partial z} \frac{\partial z}{\partial t}$$

$$\therefore \frac{D\alpha}{Dt} = \frac{\partial \alpha}{\partial t} + \mathbf{U} \text{ grad } \alpha, \text{ as } \mathbf{U} \text{ has the components } \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}.$$

For a vector \mathbf{F}

$$\frac{D\mathbf{F}}{Dt} = \frac{\partial \mathbf{F}}{\partial t} + (\mathbf{U} \text{ grad})\mathbf{F}.$$

It is well known that

$$\operatorname{div} \mathbf{J} = \operatorname{div} (\rho \mathbf{U}) + \frac{\partial \rho}{\partial t} = 0 \text{ (equation of continuity).}$$

We will assume that the force $[\mathbf{U}, \mathbf{H}]$ can be neglected. Then

$$\frac{D\mathbf{U}}{Dt} = \frac{\mu}{\rho} \mathbf{E} = \frac{e}{m} \mathbf{E}.$$

Now

$$\frac{D\mathbf{U}}{Dt} = \frac{\partial \mathbf{U}}{\partial t} + (\mathbf{U} \operatorname{grad}) \mathbf{U},$$

and as for a stationary motion

$$\frac{\partial \mathbf{U}}{\partial t} = 0, \quad \frac{\partial \rho}{\partial t} = 0$$

$$\therefore \frac{D\mathbf{U}}{Dt} = (\mathbf{U} \operatorname{grad}) \mathbf{U} = \frac{e}{m} \mathbf{E},$$

$$\operatorname{div} (\rho \mathbf{U}) = 0.$$

Using $\operatorname{div} \mathbf{E} = \rho$ this becomes

$$\operatorname{div} ((\mathbf{U} \operatorname{grad}) \mathbf{U}) = \frac{e}{m} \rho. * \text{ (Dynamical equation.)} \quad . \quad . \quad . \quad (2)$$

$$\rho \operatorname{div} \mathbf{U} + \mathbf{U} \operatorname{grad} \rho = 0. \text{ (Equation of continuity.)} \quad . \quad . \quad . \quad (3)$$

In this paper we shall make use of the following property of beams. In an actual beam (fig. 2) if we cut the outer part away (II) it will not affect the remaining inner part (I). The resultant of the forces exerted by II in region I is zero. Mathematically this will enable us to make use of an analytical continuation of the beam sideways.

$$* \quad ((\mathbf{U} \operatorname{grad}) \mathbf{U}) = \left(U_x \frac{\partial}{\partial x} + U_y \frac{\partial}{\partial y} + U_z \frac{\partial}{\partial z} \right) (U_x \mathbf{I} + U_y \mathbf{J} + U_z \mathbf{K}),$$

where $\mathbf{I}, \mathbf{J}, \mathbf{K}$ are unit vectors in direction x, y, z .

$$\begin{aligned} (\mathbf{U} \operatorname{grad}) \mathbf{U} = & \left(U_x \frac{\partial U_x}{\partial x} + U_y \frac{\partial U_x}{\partial y} + U_z \frac{\partial U_x}{\partial z} \right) \mathbf{I} \\ & + \left(U_x \frac{\partial U_y}{\partial x} + U_y \frac{\partial U_y}{\partial y} + U_z \frac{\partial U_y}{\partial z} \right) \mathbf{J} \\ & + \left(U_x \frac{\partial U_z}{\partial x} + U_y \frac{\partial U_z}{\partial y} + U_z \frac{\partial U_z}{\partial z} \right) \mathbf{K} \end{aligned}$$

$$\operatorname{div} ((\mathbf{U} \operatorname{grad}) \mathbf{U})$$

$$\begin{aligned} = & \left(\frac{\partial U_x}{\partial x} \right)^2 + U_x \frac{\partial^2 U_x}{\partial x^2} + \frac{\partial U_y}{\partial x} \frac{\partial U_x}{\partial y} + U_y \frac{\partial^2 U_x}{\partial x \partial y} + \frac{\partial U_z}{\partial x} \frac{\partial U_x}{\partial z} + U_z \frac{\partial^2 U_x}{\partial z \partial x} \\ & + \frac{\partial U_x}{\partial y} \frac{\partial U_y}{\partial x} + U_x \frac{\partial^2 U_y}{\partial y \partial x} + \left(\frac{\partial U_y}{\partial y} \right)^2 + U_y \frac{\partial^2 U_y}{\partial y^2} + \frac{\partial U_z}{\partial y} \frac{\partial U_y}{\partial z} + U_z \frac{\partial^2 U_y}{\partial y \partial z} \\ & + \frac{\partial U_x}{\partial z} \frac{\partial U_z}{\partial x} + U_x \frac{\partial^2 U_z}{\partial x \partial z} + \frac{\partial U_y}{\partial z} \frac{\partial U_z}{\partial y} + U_y \frac{\partial^2 U_z}{\partial y \partial z} + \left(\frac{\partial U_z}{\partial z} \right)^2 + U_z \frac{\partial^2 U_z}{\partial z^2}. \end{aligned}$$

It will now be advantageous to have a more special problem in mind, where some assumptions will have to be introduced. Assume the beam emerging from a rectangular slit.

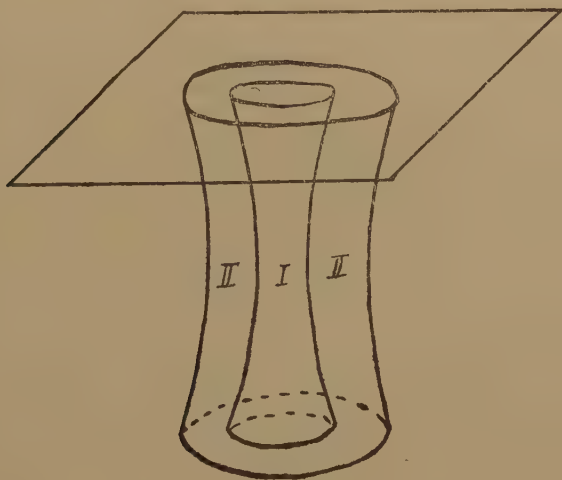
To a great degree of approximation it will be possible to write

$$\begin{aligned} \mathbf{U} &= U_x \mathbf{I} + U_y \mathbf{J} + U_z \mathbf{K} \\ &= a(z)x\mathbf{I} + b(z)y\mathbf{J} + c(z)\mathbf{K}. \end{aligned}$$

The meaning of this assumption is that in each cross-section of the beam the components U_x and U_y of the velocity are linear functions of the co-ordinates x and y respectively. For focusing purposes this may be assumed to be prearranged by the electron optical system before the entrance plane, so that for this initial plane

$$U_{x_{z=0}} = a(0)x, \quad U_{y_{z=0}} = b(0)y$$

Fig. 2.



can be regarded as given.

Since the main interest lies in converging beams where the distances from the centre of the beam become smaller and smaller, the linear relation at the entrance plane will not get distorted appreciably as we proceed along the beam, but of course the proportionality factors $a(z)$ and $b(z)$ will be functions of z . Particularly at the narrowest part of the beam all the electrons will move parallel to the z axis and $a(z_{\min})=0$, $b(z_{\min})=0$.

As to the z component of the velocity, $c(z)$, we will assume it in general to be large as compared with the other components. It will of course depend on the co-ordinate z , even if no external force is applied to accelerate or decelerate the electrons, as the space charge affects all the components. But at any rate we will neglect the changes of $c(z)$ in any given z -plane. These changes will be small if we assume $c(z)$ to be the same for all the electrons when emerging from the plane $z=0$, and of a relatively large value.

Under these assumptions, we shall have

$$\begin{aligned} (\mathbf{U} \text{ grad}) &= a(z)x \frac{\partial}{\partial x} + b(z)y \frac{\partial}{\partial y} + c(z) \frac{\partial}{\partial z}, \\ (\mathbf{U} \text{ grad})\mathbf{U} &= a^2(z)x\mathbf{I} + b^2(z)y\mathbf{J} + c(z)x \frac{da(z)}{dz}\mathbf{I} \\ &\quad + c(z)y \frac{db(z)}{dz}\mathbf{J} + c(z) \frac{dc(z)}{dz}\mathbf{K} = \frac{e}{m} \mathbf{E}. \end{aligned} \quad (5)$$

$$\begin{aligned} \text{div} ((\mathbf{U} \text{ grad})\mathbf{U}) &= a^2(z) + b^2(z) + c(z) \frac{da(z)}{dz} + c(z) \frac{db(z)}{dz} \\ &\quad + \left(\frac{dc(z)}{dz} \right)^2 + c(z) \frac{d^2c(z)}{dz^2} = \frac{e}{m} \rho. \end{aligned} \quad (6)$$

The equation of continuity (3) becomes, when using (4),

$$\rho a(z) + \rho b(z) + \rho \frac{dc(z)}{dz} + a(z)x \frac{\partial \rho}{\partial x} + b(z)y \frac{\partial \rho}{\partial y} + c(z) \frac{\partial \rho}{\partial z} = 0. \quad (7)$$

This equation can be further simplified if we assume the density $\rho(x, y, z)$ of the beam constant in each cross-section. We will of course assume that this is exactly the case at the plane of entrance. Then to the same degree of accuracy with which the other assumptions hold this uniformity of density in each cross-section will hold.

In the differential eqn. (7) $x \frac{\partial \rho}{\partial x}$ and $y \frac{\partial \rho}{\partial y}$ appear. Assume the slit to be narrow and long (x small, y big) then $\frac{\partial \rho}{\partial y} \ll \frac{\partial \rho}{\partial x}$ so that both $x \frac{\partial \rho}{\partial x}$ and $y \frac{\partial \rho}{\partial y}$ are small. The differential equations to be solved take the form:—

$$a^2(z) + b^2(z) + c(z) \frac{da(z)}{dz} + c(z) \frac{db(z)}{dz} + \left(\frac{dc(z)}{dz} \right)^2 + c(z) \frac{d^2c(z)}{dz^2} = \frac{e}{m} \rho. \quad (6)$$

$$\rho a(z) + \rho b(z) + \rho \frac{dc(z)}{dz} + c(z) \frac{d\rho}{dz} = 0. \quad (7')$$

As will be seen x and y disappear from the differential eqn. and these cease to be partial differential equations. If this had not been achieved the neglect of $\frac{\partial \rho}{\partial x}$ and $\frac{\partial \rho}{\partial y}$ would not have been justified.

THE PLANE DIODE AND A VERY LONG SLIT.

CASE 1. (Two dimensional case.)

We will first assume that the slit is very long. The case of the finite rectangular slit will be taken up again at the end of this paper. From the translational symmetry $U_y = 0$, and hence $b(z) = 0$, (6) and (7') become

$$a^2(z) + c(z) \frac{da(z)}{dz} + \left(\frac{dc(z)}{dz} \right)^2 + c(z) \frac{d^2c(z)}{dz^2} = \frac{e}{m} \rho, \quad (8)$$

$$\rho a(z) + \rho \frac{dc(z)}{dz} + c(z) \frac{d\rho}{dz} = 0. \quad (9)$$

Before we deal with these equations it will be interesting to assume the slit as infinitely wide, with electrons having an entrance velocity only along the z axis. Then from the translational symmetry $U_x=0$ throughout and $a(z)=0$. All the previous assumptions will hold exactly, and we have the problem of the *plane diode* :

$$\left(\frac{dc(z)}{dz}\right)^2 + c(z)\frac{d^2c(z)}{dz^2} = \frac{e}{m}\rho, \quad \dots \dots \dots (8')$$

$$\rho\frac{dc(z)}{dz} + c(z)\frac{d\rho}{dz} = 0. \quad \dots \dots \dots (9')$$

Equation (9') can be integrated at once

$$\frac{d(\rho c(z))}{dz} = 0, \quad \rho c(z) = \frac{1}{k} = \text{constant}, \quad \dots \dots \dots (10)$$

$$\frac{1}{k} = c(z)\rho = J = \text{the current density.}$$

From (10)

$$\frac{dc(z)}{dz} = -\frac{1}{k\rho^2}\frac{d\rho}{dz}, \quad \dots \dots \dots (11)$$

$$\frac{d^2c(z)}{dz^2} = \frac{2}{k\rho^3}\left(\frac{d\rho}{dz}\right)^2 - \frac{1}{k\rho^2}\frac{d^2\rho}{dz^2}. \quad \dots \dots \dots (12)$$

Inserting (11) and (12) in (8'), we get

$$-\rho\frac{d^2\rho}{dz^2} + 3\left(\frac{d\rho}{dz}\right)^2 = \frac{e}{m}k^2\rho^5. \quad \dots \dots \dots (13)$$

(The differential equation of the plane diode.)

For this differential equation to be solved the boundary conditions must be added

$$\rho_{z=0} = \rho_0, \quad \left(\frac{d\rho}{dz}\right)_{z=0} = \rho'_0 = -k\rho_0^2\left(\frac{dc(z)}{dz}\right)_{z=0}$$

(as follows from (11)).

ρ_0 is given. $\left(\frac{d\rho}{dz}\right)_{z=0}$ or $\left(\frac{dc(z)}{dz}\right)_{z=0}$ is not given. (k is given $k = \rho_0 c(z)_{z=0}$.)

This introduces a difficulty which we will discuss further on.

A *special solution* of (13) can be found easily by trying $\rho = pz^n$. Introducing this in (13) gives

$$2n-2=5n, \quad n = -\frac{2}{3}, \quad \rho = pz^{-2/3},$$

and

$$-\frac{10}{9}p^{2z^{-10/3}} + \frac{12}{9}p^{2z^{-10/3}} = \frac{e}{m}k^2p^{5z^{-10/3}}$$

$$\therefore p = \sqrt[3]{\frac{2}{9}\frac{m}{ek^2}}$$

$$\rho = \sqrt[3]{\frac{2}{9}\frac{m}{e}\frac{J^2}{z^2}}.$$

This is in agreement with Child's law, which corresponds to

$$\rho_0 = \infty, \quad \left(\frac{d\rho}{dz}\right)_{z=0} = -\infty \quad \text{and} \quad c(z)_{z=0} = 0, \quad \left(\frac{dc(z)}{dz}\right)_{z=0} = 0.$$

(These boundary conditions mean an infinite supply of electrons with no initial velocity. To obtain Child's law it is necessary to make these assumptions.)

But it is also possible to give the general solution of (13). Write

$$\rho = y^n(z),$$

$$\frac{d\rho}{dz} = ny^{n-1}(z)y'(z),$$

$$\frac{d^2\rho}{dz^2} = n(n-1)y^{n-2}(z)y'^2(z) + ny^{n-1}(z)y''(z),$$

and (13) becomes

$$-n(n-1)y^{2n-2}y'^2 - ny^{2n-1}y'' + 3n^2y^{2n-2}y'^2 = Ay^{5n},$$

where

$$A = \frac{e}{m} k^2.$$

If we choose n so that

$$-n(n-1) + 3n^2 = 0 \quad \therefore \quad n = -\frac{1}{2}, \quad \rho = \frac{1}{\sqrt{y(z)}},$$

the differential equation (13) becomes

$$y'' = 2Ay^{-1/2} \quad \dots \dots \dots (14)$$

$$\therefore \quad \frac{dy}{dz} = \sqrt{8A\sqrt{y} + B} \quad (B, \text{ an integration constant}) \quad \dots \dots (15)$$

$$\frac{dy}{dz} = -\frac{1}{\rho^2} \frac{d\rho}{dz} = \sqrt{\frac{8A}{\rho} + B}, \quad -\frac{1}{\rho_0^2} \left(\frac{d\rho}{dz}\right)_{z=0} = \sqrt{\frac{8A}{\rho_0} + B}. \quad \dots (16)$$

A further integration gives

$$z + D = \frac{1}{32A^2} \left[\frac{2}{3} \left(\frac{8A}{\rho} + B \right)^{3/2} - 2B \left(\frac{8A}{\rho} + B \right)^{1/2} \right], \quad \dots \dots (17)$$

where $A = \frac{e}{m} k^2, \quad \frac{1}{k} = \rho(z)c(z) = \rho_0 c_0, \quad \dots \dots \dots (18)$

(17) is the general solution of (13) but it contains two constants which cannot be determined from the actual initial conditions ($\rho_{z=0} = \rho_0, c_{z=0} = c_0$). Therefore this general integral of (13) does not yield us the solution of the plane diode with initial conditions which differ from those made in deriving Child's law. It amounts to the necessity of determining the force at $z=0$. From (5) it will be seen that

$$c(z) \frac{dc(z)}{dz} = \frac{e}{m} E_z,$$

so that if $E_{z(z=0)}$ was known, we should have

$$\left(\frac{dc(z)}{dz}\right)_{z=0} = \frac{e}{m} \frac{E_{z(z=0)}}{c_0},$$

and
$$\left(\frac{d\rho}{dz}\right)_{z=0} = k\rho_0^2 \left(\frac{dc(z)}{dz}\right)_{z=0} = k\rho_0^2 \frac{e}{m} \frac{E_{z(z=0)}}{c_0}.$$

This impossibility of determining the constants of integration is a result of the very nature of the method, which consists in dealing not with the original Maxwell-Lorentz equations but with a single differential equation which is obtained by eliminating all but one of the quantities and which is of one order higher than the original equations.

Actually the whole value of this method of eliminating all but one of the quantities, ρ , will make itself felt only, if for other physical reasons we can assume $\left(\frac{d\rho}{dz}\right)_{z=0}$ or $\left(\frac{dc(z)}{dz}\right)_{z=0}$ to be known. The more accurate our knowledge about these quantities is, the more valuable will be the method followed here.

In the electron beams which we are considering the velocity $c(z)$ is comparatively large, and we may assume that the changes which take place in this quantity can be neglected, so that

$$\frac{dc(z)}{dz} = 0, \quad c(z) = \text{const.} = c.$$

By making this assumption we are no longer confronted with the difficulty which we encountered when dealing with the diode.

As $c(z) = c = \text{constant}$, the width of the beam will be inversely proportional to the density ρ .

We will now return to the problem of the very long slit. (8) and (9) become

$$a^2(z) + c \frac{da(z)}{dz} = \frac{e}{m} \rho. \quad (19)$$

$$\rho a(z) + c \frac{d\rho}{dz} = 0. \quad (20)$$

From (20)
$$a(z) = -\frac{c}{\rho} \frac{d\rho}{dz}, \quad (21)$$

$$\frac{d\rho}{dz} a(z) + \rho \frac{da(z)}{dz} + c \frac{d^2\rho}{dz^2} = 0, \quad (22)$$

$$\rho \frac{da(z)}{dz} = -c \frac{d^2\rho}{dz^2} + \frac{c}{\rho} \left(\frac{d\rho}{dz}\right)^2. \quad (23)$$

Inserting (23) in (19) we get

$$-\rho \frac{d^2\rho}{dz^2} + 2 \left(\frac{d\rho}{dz}\right)^2 = \frac{e}{m} \frac{\rho^3}{c^2}. \quad (24)$$

$$s = \frac{AI}{2c}z^2 + \frac{a_0 s_0}{c}z + s_0;$$

$$s = \frac{eI}{2mc^3}z^2 + \frac{a_0 s_0}{c}z + s_0. \quad (30)$$

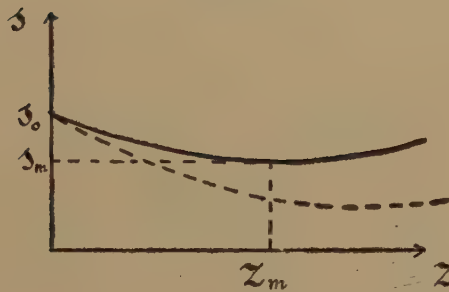
The minimum s will occur at z_{\min} .

$$\frac{eI}{mc^3}z_{\min} + \frac{a_0 s_0}{c} = 0,$$

$$z_{\min} = -\frac{a_0 s_0 mc^2}{Ie} \left(= -\frac{a_0 mc}{\rho_0 e} \right),$$

$$s_{\min} = s_0 - \frac{a_0^2 s_0^2 mc}{2eI}.$$

Fig. 3.



Full curve $|a_0|$ small; dotted curve $|a_0|$ large.

The bigger $|a_0|$ the narrower can the beam become and this will happen at correspondingly greater distances from the origin.

The result obtained here for the very long slit is exactly the same as was obtained by B. J. Thompson and F. L. Headrick by a totally different method (Proc. I.R.E. July 1940, p. 318). From (30) we can write

$$z = \frac{2mc^2}{Ie} \left[\sqrt{\frac{s_0 a_0^2}{4} - \frac{eI(s_0 - s)}{2mc}} - \frac{s_0 a_0}{2} \right]. \quad (31)$$

In the notation of Thompson and Headrick

$$z = d, \quad c = V_d, \quad \frac{s_0 a_0}{2} = V_y, \quad \frac{s_0 - s}{2} = Y_0 - y,$$

so that (31) becomes

$$d = \frac{2V_d^2 m}{Ie} \left[\sqrt{V_y^2 - \frac{Ie}{V_d m} (Y_0 - y)} - V_y \right].$$

This is the formula given by the above authors (*loc. cit.*). Therefore a further discussion of this result is unnecessary and the reader may be referred to the above paper.

CASE 2. *The Circular Beam.*

Under the same kind of general assumptions as were made for a slit we shall have

$$\begin{aligned} U &= U_r R + U_\phi \Phi + U_z K \\ (4) \text{ becomes in this case } U &= a(z)rR + c(z)K, \\ U_r &= a(z)r, \quad U_z = c(z), \quad U_\phi = 0 \end{aligned} \quad (32)$$

To find the expression for $(U \text{ grad})$ in cylindrical co-ordinates, we must express the scalar product in cylindrical co-ordinates.

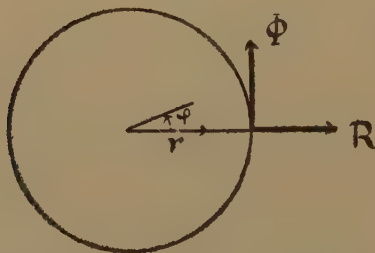
$$(AB) = A_x B_x + A_y B_y + A_z B_z = A_r B_r + A_\phi B_\phi + A_z B_z.$$

(This would apply to any orthogonal system of co-ordinates.) The components of grad in cylindrical co-ordinates are

$$\frac{\partial}{\partial r}, \quad \frac{1}{r} \frac{\partial}{\partial \phi}, \quad \frac{\partial}{\partial z}.$$

Therefore
$$(U \text{ grad}) = U_r \frac{\partial}{\partial r} + U_\phi \frac{1}{r} \frac{\partial}{\partial \phi} + U_z \frac{\partial}{\partial z}.$$

Fig. 4.



R, Φ, K unit-vectors in direction r, ϕ, z , the cylindrical co-ordinates.

In our case

$$(U \text{ grad}) = a(z)r \frac{\partial}{\partial r} + c(z) \frac{\partial}{\partial z},$$

$$(U \text{ grad})U = a^2(z)rR + c(z)r \frac{da(z)}{dz} R + c(z) \frac{dc(z)}{dz} K,$$

$$\text{div} ((U \text{ grad})U) = \frac{1}{r} \frac{\partial}{\partial r} \left[a^2(z)r^2 + r^2 c(z) \frac{da(z)}{dz} \right] + \frac{\partial}{\partial z} \left[c(z) \frac{dc(z)}{dz} \right] = \frac{e}{m} \rho.$$

$$2a^2(z) + 2c(z) \frac{da(z)}{dz} + \left(\frac{dc(z)}{dz} \right)^2 + c(z) \frac{d^2 c(z)}{dz^2} = \frac{e}{m} \rho. \quad (33)$$

The equation of continuity in cylindrical co-ordinates becomes

$$\begin{aligned} \rho \text{ div } U + U \text{ grad } \rho &= \rho \left[\frac{1}{r} \frac{\partial}{\partial r} (rU_r) + \frac{1}{r} \frac{\partial U_\phi}{\partial \phi} + \frac{\partial U_z}{\partial z} \right] \\ &\quad + U_r \frac{\partial \rho}{\partial r} + U_\phi \frac{1}{r} \frac{\partial \rho}{\partial \phi} + U_z \frac{\partial \rho}{\partial z}. \end{aligned}$$

Using (32) this becomes

$$2\rho a(z) + \rho \frac{dc(z)}{dz} + ra(z) \frac{\partial \rho}{\partial r} + c(z) \frac{\partial \rho}{\partial z} = 0. \quad (34)$$

Again, if we assume $\frac{\partial \rho}{\partial r} = 0$, and $c(z) = \text{const.} = c$ (33) and (34) become

$$2a^2(z) + 2c \frac{da(z)}{dz} = \frac{e}{m} \rho, \quad (35)$$

$$2\rho a(z) + c \frac{d\rho}{dz} = 0. \quad (36)$$

From (36)

$$a(z) = -\frac{c}{2\rho} \frac{d\rho}{dz}, \quad (37)$$

and

$$2 \frac{d\rho}{dz} a(z) + 2\rho \frac{da(z)}{dz} + c \frac{d^2 \rho}{dz^2} = 0,$$

$$\frac{da(z)}{dz} = -\frac{c}{2\rho} \frac{d^2 \rho}{dz^2} + \frac{c}{2\rho^2} \left(\frac{d\rho}{dz} \right)^2. \quad (38)$$

Inserting (37) and (38) in (35) we get

$$\begin{aligned} \frac{2c^2}{4\rho^2} \left(\frac{d\rho}{dz} \right)^2 - \frac{2c^2}{2\rho} \frac{d^2 \rho}{dz^2} + \frac{c^2}{\rho^2} \left(\frac{d\rho}{dz} \right)^2 &= \frac{e}{m} \rho. \\ -\rho \frac{d^2 \rho}{dz^2} + \frac{3}{2} \left(\frac{d\rho}{dz} \right)^2 &= \frac{e}{mc^2} \rho^3, \end{aligned} \quad (39)$$

$$\rho(z)_{z=0} = \rho_0, \quad \left(\frac{d\rho}{dz} \right)_{z=0} = -\frac{2\rho_0 \alpha_0}{c}. \quad (40)$$

(as follows from (37)).

Let $\rho = y^n(z)$. Inserting this in (39) and making $3/2n^2 - n(n-1) = 0$

we get
$$n = -2, \quad \rho = \frac{1}{y^2(z)},$$

$$2y''(z) = \frac{A}{y}, \quad (41)$$

$$\frac{dy}{dz} = \sqrt{A \ln y + B} \quad (B = \text{constant}), \quad (42)$$

$$z + D = \int \frac{dy}{\sqrt{A \ln y + B}} \quad (D = \text{constant}). \quad (43)$$

Substituting $y = e^{q^2 - B/A}$, (43) becomes

$$z + D = \frac{2e^{-B/A}}{\sqrt{A}} \int e^{q^2} dq.$$

The value of the constant B can be determined from (42). As

$$y = \frac{1}{\rho^{1/2}},$$

$$\frac{dy}{dz} = -\frac{1}{2} \rho^{-3/2} \frac{d\rho}{dz},$$

$$\left(\frac{dy}{dz}\right)^2 = \frac{1}{4} \rho^{-3} \left(\frac{d\rho}{dz}\right)^2.$$

Using (42) :

$$\frac{1}{4} \frac{1}{\rho_0^3} \left(\frac{d\rho}{dz}\right)^2_{z=0} + \frac{A}{2} \ln \rho_0 = B.$$

Using (40) this becomes

$$B = \frac{a_0^2}{\rho_0 c^2} + \frac{A}{2} \ln \rho_0.$$

The final result will be

$$z = \frac{2e^{-B/A}}{\sqrt{A}} \int_{q_0}^q e^{q^2} dq, \dots \dots \dots (44)$$

where

$$q = \sqrt{\frac{B - \frac{A}{2} \ln \rho}{A}}, \dots \dots \dots (45)$$

$$q_0 = \sqrt{\frac{B - \frac{A}{2} \ln \rho_0}{A}} = \sqrt{\frac{a_0^2}{\rho_0^2 c^2}} = a_0 \sqrt{\frac{m}{\rho_0 e}} \dots \dots \dots (46)$$

This is again identical with the result obtained by Thompson and Headrick.

CASE 3. *A Rectangular Slit of Finite Length.*

We will assume that the beam is fired with velocity gradients a_0 and b_0

$$a_0 = \epsilon b_0$$

where ϵ is a constant.

Again, we may assume that the beam will retain its uniform density in each cross-section so that $\frac{\partial \rho}{\partial x} = \frac{\partial \rho}{\partial y} = 0$, and as before $c(z) = c = \text{constant}$.

(6) and (7) become

$$a^2(z) + b^2(z) + c \frac{da(z)}{dz} + c \frac{db(z)}{dz} = \frac{e}{m} \rho, \dots \dots \dots (47)$$

$$\rho a(z) + \rho b(z) + c \frac{d\rho}{dz} = 0. \dots \dots \dots (48)$$

In these two equations we have three unknowns ; $a(z)$, $b(z)$ and $\rho(z)$, and therefore we need another relation. This can be found easily for $a(z)$ and $b(z)$.

$$\begin{aligned} \text{Curl curl } H &= \text{curl } J = \text{curl } (\rho U) = 0 \text{ (as } H \cong 0, \dot{E} = 0). \\ \text{Curl } (\rho U) &= \rho \text{ curl } U + [\text{grad } \rho, U] \end{aligned}$$

$$= \rho \left| \begin{array}{ccc} \text{I} & \text{J} & \text{K} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a(z)x & b(z)y & c(z) \end{array} \right| + \left| \begin{array}{ccc} \text{I} & \text{J} & \text{K} \\ \frac{\partial \rho}{\partial x} & \frac{\partial \rho}{\partial y} & \frac{\partial \rho}{\partial z} \\ a(z)x & b(z)y & c(z) \end{array} \right| = 0.$$

As $\frac{\partial \rho}{\partial x} = 0$ and $\frac{\partial \rho}{\partial y} = 0$, we get from the I component

$$-\rho y \frac{db(z)}{dz} - \frac{d\rho}{dz} b(z)y = 0 \quad (49)$$

From the J component we get

$$\rho x \frac{da(z)}{dz} + a(z)x \frac{d\rho}{dz} = 0 \quad (50)$$

(The K component is zero.)

From (49) and (50) follows

$$\frac{1}{b(z)} \frac{db(z)}{dz} = \frac{1}{a(z)} \frac{da(z)}{dz}$$

$\therefore b(z) = ka(z)$ where k is a constant.

Hence we get the useful result that the ratio of the gradients of the velocities in the x and y direction remains constant.

Therefore if

$a_0 = b_0$, we shall have

$$a(z) = b(z) \quad (51)$$

Inserting (51) in (47) and (48) we get :

$$a^2(z) \left(1 + \frac{1}{\epsilon^2}\right) + c \left(1 + \frac{1}{\epsilon}\right) \frac{da(z)}{dz} = \frac{e}{m} \rho \quad (52)$$

$$\rho a(z) \left(1 + \frac{1}{\epsilon}\right) + c \frac{d\rho}{dz} = 0 \quad (53)$$

From (53)

$$a(z) = -\frac{c}{\rho} \frac{d\rho}{dz} \left(1 + \frac{1}{\epsilon}\right)^{-1} \quad (54)$$

and

$$\left(1 + \frac{1}{\epsilon}\right) \frac{d\rho}{dz} a(z) + \left(1 + \frac{1}{\epsilon}\right) \rho \frac{da(z)}{dz} + c \frac{d^2\rho}{dz^2} = 0,$$

$$\rho \frac{da(z)}{dz} = -c \left(1 + \frac{1}{\epsilon}\right)^{-1} \frac{d^2\rho}{dz^2} - a(z) \frac{d\rho}{dz},$$

$$\left(1 + \frac{1}{\epsilon}\right) \frac{da(z)}{dz} = -\frac{c}{\rho} \frac{d^2\rho}{dz^2} + \frac{c}{\rho^2} \left(\frac{d\rho}{dz}\right)^2.$$

Inserting this in (52) and using (54) we get

$$\frac{\left(1 + \frac{1}{\epsilon^2}\right) + \left(1 + \frac{1}{\epsilon}\right)^2}{\left(1 + \frac{1}{\epsilon}\right)^2} \left(\frac{d\rho}{dz}\right)^2 - \rho \frac{d^2\rho}{dz^2} = \frac{e}{mc^2} \rho^3 \quad (55)$$

The boundary conditions are

$$\rho(z)_{z=0} = \rho_0, \quad \left(\frac{d\rho}{dz}\right)_{z=0} = -\frac{a_0 \left(1 + \frac{1}{\epsilon}\right) \rho_0}{c} \quad (\text{as follows from (54)}). \quad (56)$$

Since interchanging width and height should make no difference the coefficient of $\left(\frac{d\rho}{dz}\right)^2$ in (55) should not alter its value if we replace ϵ by $1/\epsilon$. This is indeed the case, so that (55) can also be written as

$$\frac{(1+\epsilon^2)+(1+\epsilon)^2}{(1+\epsilon)^2} \left(\frac{d\rho}{dz}\right)^2 \rho - \frac{d^2\rho}{dz^2} = \frac{e}{mc^2} \rho^3. \quad (57)$$

If we replace $1/\epsilon$ by ϵ we must replace a by b in the boundary condition (56).

For $\epsilon=1$ which would correspond to focusing to a point we get the same differential equation as was obtained for a circle, with the same boundary conditions. This is not surprising, since we can take away outside layers from such a beam and leave only a circular part, which is not affected by outside layers.

The coefficient

$$\frac{(1+\epsilon^2)+(1+\epsilon)^2}{(1+\epsilon)^2}$$

varies from $3/2$ to 2 if

$$1 \leq \epsilon < \infty.$$

(57) can be integrated in the manner shown in this paper for any intermediate value of ϵ . If ϵ is different from 1 the beam is originally not focused to a point but to two lines, not in the same plane.

Bibliography.

- (1) E. E. Watson, "Dispersion of an Electron Beam," *Phil. Mag.* 7th Series, vol. iii. p. 849 (1927).
- (2) B. J. Thompson and L. B. Headrick, "Space Charge Limitations in the Focus of Electron Beams," *Proc. I. R. E.* vol. xxviii. nr. 7, p. 318 (July 1940).

VI. Note on the Solubility of Hydrogen in Palladium.

By Professor E. A. OWEN, M.A., Sc.D.,
University College of North Wales, Bangor*.

[Received December 3, 1943.]

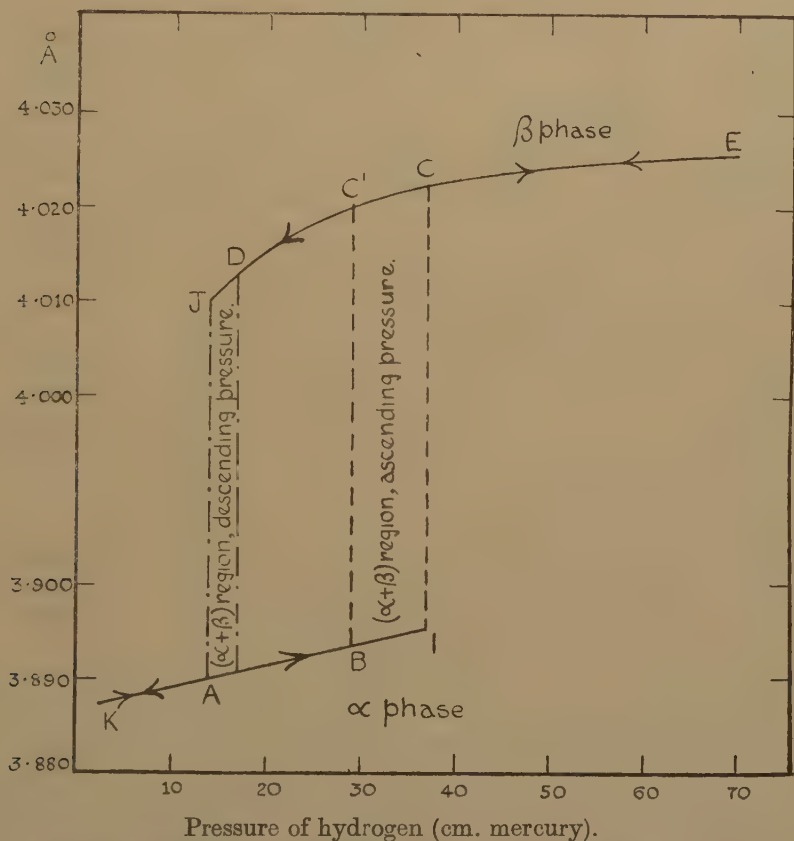
1. Introduction.

A GOOD deal of attention has been directed to the occlusion of hydrogen by palladium, and in recent years the system has been investigated by X-ray methods with some interesting results⁽¹⁻⁷⁾. A valuable con-

* Communicated by the Author.

tribution to the investigation was made by Lacher⁽⁸⁾ who, after making certain assumptions, derived theoretical formulæ for the solubility of hydrogen in palladium and who, in the course of his paper, made some interesting suggestions regarding the existence of the "hysteresis" effect observed when the pressure is taken through a cycle of increasing and decreasing values. For some years before war broke out, the problem had been examined in this laboratory, by X-ray methods, over a range of temperature which, though limited in comparison with that

Fig. 1.



Relation between parameter and pressure of hydrogen at 100° C.
in the palladium-hydrogen system.

of other workers employing a different method of investigation, yet provided information which went a long way to confirm the ideas put forward by Lacher. It is the purpose of this note to supplement what has already been published on the problem from this laboratory and to consider the experimental results as far as they go in the light of Lacher's theoretical treatment.

The X-ray method is very suitable for the investigation of the solution of hydrogen in palladium, because information may be obtained not only of the total effect but of the individual behaviour of the two phases present in the system, as the pressure of the gas is varied. A detailed account of observations made within narrow limits of temperature and pressure has already been recorded ^{(5), (6), (7)}. It is found that when the pressure is increased the two phases exist together over a certain range of pressure, but that they exist over a narrower range and at a lower mean value of the pressure when the pressure is decreased (see fig. 1). This behaviour of the system when hydrogen is added and removed in a cyclic manner, yields an effect very similar to that of hysteresis found in other fields of observation. The phenomenon is somewhat more complicated than that recorded by Gillespie and his co-workers ^{(3), (4)} who, by using special preparations of palladium, were able to trace the isothermal curves in a reversible manner and therefore without evidence of hysteresis. With ordinary palladium metal, however, the "hysteresis" effect appears always to be present; the effect is quite definite and repeatable under various conditions of temperature and pressure.

2. Pressure—Concentration Isothermals.

It is usual to express the concentration of the solution by the ratio of the number of hydrogen to the number of palladium atoms (r , say), and to denote the two phases of the system by α and β , the α being the palladium-rich and the β the hydrogen-rich phase. The only measurements made by the X-ray method in this laboratory were those on the parameters of the lattices that are found in the system. The curves obtained at 100° C., when the pressure of the gas was kept steady for two hours before the exposure was made, are shown in fig. 1 (curves KI and JE); they give the relations existing between the pressure of the gas and the parameters of the lattices.

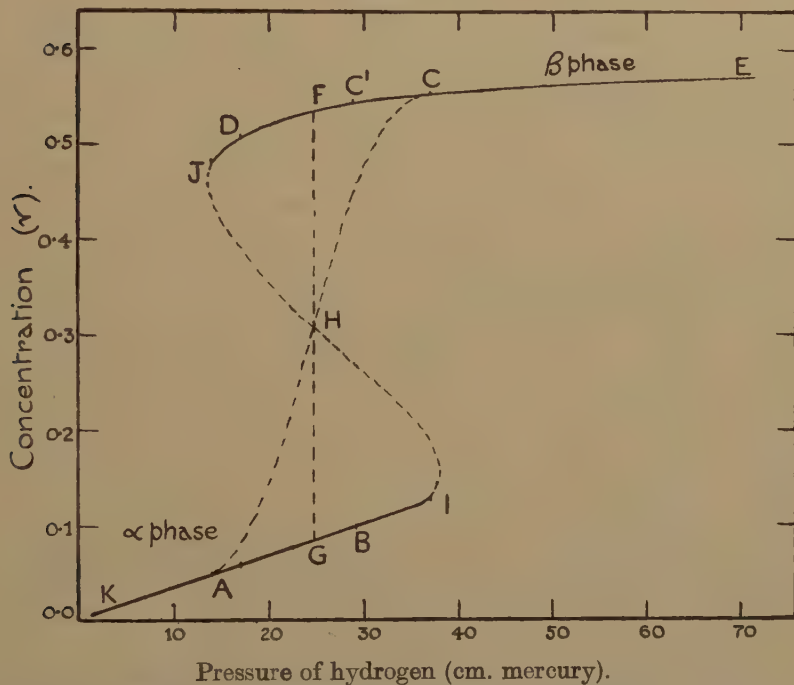
No direct measurement of the concentration was made; this was determined, for any combination of temperature and pressure, directly or by extrapolation from the pressure-concentration isothermals of Hoitsema and Roozeboom ⁽⁹⁾. The concentration of the individual phases in two-phase regions had always to be calculated by extrapolation, and this was done by finding the rate of change of the parameter with pressure in those portions of the isothermals where only one constituent was present, the change of parameter with pressure being approximately linear in these portions. The resulting curves (KI and JE) at 100° C. are shown in fig. 2.

As hydrogen enters the palladium, the concentration increases approximately linearly with the pressure except possibly at very low pressures. When the pressure is gradually increased from a very low value it is found, at 100° C., that the β -phase makes its first appearance at B (or C') when the pressure is about 29 cm. of mercury and $r=0.10$. Beyond

B, as the pressure is further increased, the α -phase gradually disappears and the amount of the β -phase increases until at C (or I), when the pressure is about 37 cm. of mercury, the system consists entirely of β -phase with $r=0.55$. As the pressure is still further increased to atmospheric the parameter of the β -lattice gradually increases and there is a slight increase of concentration.

It is interesting to note that the parameter of the α -lattice increases in the region BI, where the two phases co-exist, so that the α -phase, although gradually decreasing in amount, takes in hydrogen in crossing

Fig. 2.



Relation between concentration and pressure of hydrogen at 100° C. in the palladium-hydrogen system.

the two-phase region, with the result that it is richer in hydrogen as it nears the point where it disappears than it was when the β -phase first made its appearance. Similarly, the concentration of the β -phase changes in crossing this two-phase region. This is contrary to what normally happens in two-phase regions of alloy systems in thermal equilibrium; in those cases the concentration of each phase remains constant throughout a two-phase region. It would appear, therefore, that the system is not in thermal equilibrium.

With increasing pressure the curves KI and C'E are obtained, the former referring to the α -phase and the latter to the β -phase. When the

pressure is lowered from atmospheric, the β -phase persists until the point D is reached, which is well removed from the point C where the α -phase disappeared when the pressure was increased. Thus, when hydrogen is withdrawn from the β -phase, the latter exists alone until r reaches the value 0.51, whereas when the pressure was increasing the α -phase did not disappear until r was 0.55. By still lowering the pressure, the β -phase rapidly decreases in amount and also in concentration across the two-phase region, until at J it has entirely disappeared and the system again consists of pure α -phase. At A, where the β -phase disappears with decreasing pressure, r has the value 0.05, this being only half its value when the β -phase first appeared with increasing pressure. After this stage, on further withdrawal of hydrogen, the curve AK, which was obtained with increasing pressure, is followed. This cycle of changes can be repeated if the material is annealed to bring it to its initial state after each cycle of operations.

The Hoitsema and Roozeboom isothermal at 100°C . is the curve KAHCE, the middle curved portion AHC covering the region where the α - and β -phases appear together either on increasing or decreasing the pressure.

Assuming that it is possible to produce a "supersaturated" α -phase and an "undersaturated" β -phase, as suggested by Lacher, then a curve such as KIHJE can be drawn to represent the changes that take place, the portion GIH representing supersaturated α -solutions and the portion FJH, undersaturated β -solutions. No points on the hypothetical portion JHI of this curve were observed. If these states of metastable equilibrium did not exist, the saturation curve would be represented by a curve such as KGHFE, where FG is situated half-way between the mean positions of the two-phase regions. At G, $r_\alpha=0.08$ and at F, $r_\beta=0.53$, so that $r_s=r_\alpha+r_\beta=0.61$, a value which is in good agreement with the value derived from the observations of Gillespie and Hall in the region of 100°C . The values of r derived similarly from the 80°C . and 120°C . isothermals are 0.63 and 0.59 respectively. The mean value for these three temperatures is therefore 0.61, which is near the average value of 0.59 taken for r_s by Lacher in his calculations. This quantity is found to be fairly constant over a wide range of temperature which, Lacher points out, indicates that the number of holes in palladium available for hydrogen is constant and closely equal to the number of holes in the d band of palladium. Hence it is probable that the hydrogen in solution is in a completely ionized state and that the electrons go into the d shell of the palladium atom. It follows from the above that palladium becomes saturated with hydrogen when r is about 0.6.

3. Diffuseness of the Lines in the Structure-spectra.

In the initial stages of the investigation of the palladium-hydrogen system, the lack of definition of the lines in the structure-spectra was

believed to indicate a distortion of the crystal lattice, but it became evident from the study of the variation in the width of the lines with glancing angle that the lack of definition was due not to distortion but to small particle size.

Starting with annealed palladium, the structure-spectrum of which has sharply defined lines, the α -phase lattice increases in dimensions as hydrogen enters the material, and its lines remain sharply defined even after the appearance of the β -phase. As the pressure of the gas is increased, the α -phase lines become weaker but still remain sharp in outline, whilst those of the β -phase become stronger and from their first appearance are diffuse. On further increase of pressure, and again on decreasing the pressure over the regions where the α -phase is absent, the β -phase lines remain badly defined. When the α -phase makes its reappearance on further decrease of pressure, the lines in its structure-spectrum are now also diffuse. A heat treatment at 250° C. was not drastic enough to improve, in reasonable time, the definition of the lines; it required annealing at 450° C. for some hours to produce the desired effect.

Since the definition of the β -phase lines remains poor as more hydrogen is absorbed, and there is no improvement on further annealing after the α -phase has disappeared, it would appear that the β -phase is formed in small nuclei. When the α -phase makes its reappearance on diminishing the pressure, it is also formed in nuclei and remains as a conglomeration of small particles after the whole of the system has changed over into the α -condition, these particles being small enough to scatter the X-rays, thus causing diffuseness of the lines.

The improved apparatus⁽⁷⁾ used in the later investigations was well suited to examine the effect on the definition of the lines of prolonged charging of the specimen with hydrogen at constant temperature. When the β -phase was present alone, several days annealing in a hydrogen atmosphere at the temperature of the isothermal failed to produce lines that were sharp in outline. The experimental procedure was to mount the palladium specimen in a hydrogen atmosphere, the temperature and pressure of which were kept constant over a period of five days, and to take as many X-ray photographs as possible during that period. Each series of photographs taken at any given temperature and pressure, provided evidence of the diffuseness of the β -phase lines and also of the α -phase lines when this phase is formed from the β -phase.

It is evident that, at the low temperatures employed, no large crystals of solid solution are formed during the transformations that take place; the only phase containing large crystals is the initial α solid solution.

4. Conclusions.

It may be concluded from the above observations that the β -phase when first formed is finely dispersed throughout the α -phase, and that

it remains finely divided even when the whole material consists of β -phase. Likewise the α -phase, when it reappears, is finely dispersed throughout the β -phase, and when all the parent β -phase has disappeared, the newly formed α -phase remains in a finely divided state unless it is annealed for some time at a temperature substantially higher than that of the isothermal. These are precisely the assumptions made by Lacher, on the basis of which he tentatively derives a qualitative explanation of the hysteresis loops.

When once the material is brought to a condition in which it consists of fine particles, it appears to be difficult to get these particles to coalesce and produce large crystals of material that will yield photographs with well-defined lines. In periods of time extending up to five days, when the material was kept at the temperature of the isothermal, no improvement was observed in the definition of the lines. To produce an improvement in reasonable time the temperature had to be raised. These facts point to a sluggishness in the behaviour of the system that might well account for the hysteresis effect observed. In some respects the system resembles the iron-nickel system, in which irreversible changes are observed with certain alloys rich in iron, and in this region the system displays the same reluctance to change that is shown by the palladium-hydrogen system. Under these conditions the systems are not in true thermal equilibrium; they are in metastable states which are very persistent.

It is to be noted that the hysteresis loops increase in width and in area as the temperature is raised. This is contrary to expectation if, as it is reasonable to assume, the system attains its state of equilibrium more readily the higher the temperature. To reconcile the observations with Lacher's hypothesis that there exist supersaturated and under-saturated solutions in the palladium-hydrogen system, it would be necessary to assume that the range of metastability increases with rise of temperature, which does not appear probable.

No evidence was obtained in the course of the investigations that metallic hydrides are formed within the limited range of temperature employed; all the experimental results point to two solid solutions. This result, and the experimental evidence that the system is not in equilibrium, confirm the two main assumptions made by Lacher upon which he bases his theoretical treatment of the solubility of hydrogen in palladium. Lacher's further assumption, that the phases in the system exist in finely divided states, is borne out by experiment.

It is to be emphasized that the present experimental results are confined to pressures below one atmosphere and to a narrow range of temperature from 60° C. to 130° C. At greater pressures and higher temperatures the results may be different as, in fact, the work of Gillespie and his colleagues definitely indicates. But even at the low temperatures and pressures employed in the present work, it would appear that further consideration will have to be given to the hysteresis phenomenon to explain it in all its details.

References.

- (1) Linde, J. O., and Borelius, G., *Ann. d. Physik*, lxxxiv. p. 747 (1927).
- (2) Sieverts, A., and Zaff, G., *Z. phys. Chem. A*, clxiv. p. 359 (1935).
- (3) Gillespie, L. J., and Hall, F. P., *J. Amer. Chem. Soc.* xlviii. p. 1207 (1926).
- (4) Gillespie, L. J., and Galstaun, L. S., *J. Amer. Chem. Soc.* lviii. p. 2565 (1936).
- (5) Owen, E. A., and Jones, J. I., *Proc. Phys. Soc.* xlix. p. 587 (1937).
- (6) Owen, E. A., and Jones, J. I., *Proc. Phys. Soc.* xlix. p. 603 (1937).
- (7) Owen, E. A., and Williams, E. St. J., *Proc. Phys. Soc.* lvi. p. 52 (1944).
- (8) Lacher, L. R., *Proc. Roy. Soc. A*, clxi. p. 525 (1937).
- (9) Hoitsema and Roozeboom, *Z. phys. Chem.* xvii. p. 1 (1895).

VII. *Viscosities at the Boiling Point of some Primary Amines,
Cyclohexane and some of its Derivatives.*

By J. NEWTON FRIEND and WILLIAM D. HARGREAVES*.

[Received October 27, 1943.]

THE rheochor, R , has already⁽¹⁾ been defined as $M(10^3\eta)^{1/8}/(D+2d)$, where M is the molar weight, D and d are the density of liquid and vapour, and η the viscosity. The main objects of this research have been to evaluate the rheochor of the amino-group and to find the effect of six-carbon ring formation. The following atomic, group and structural rheochors have already been evaluated at the boiling point:—C=12.8; H (attached to C), 5.5; OH (hydroxyl), 20; O (etheric), 10; Cl, 27.3; Br, 35.8; I, 47.6; NO₂, 38.9; C₆H₅ (attached to alkyl), 100.7; \rightarrow (co-ordinate bond), -0.4.

Primary Amines.

Viscosity-temperature curves for the aromatic amines are shown in fig. 1. The great effect of the NH₂-group is seen by comparing aniline with benzene⁽²⁾, benzylamine with toluene⁽²⁾, chlor-aniline with chlor-benzene, *o*-anisidine with anisole, and (in fig. 2) cyclohexylamine with cyclohexane. The methyl group in toluene and the meta- and para-toluidines has caused a reduction in viscosity.

The toluidines illustrate the rule that the physical properties of ortho-substituted benzene derivatives are frequently abnormal when compared with those of their meta- and para-isomerides. The viscosity of ortho-toluidine lies close to that of aniline.

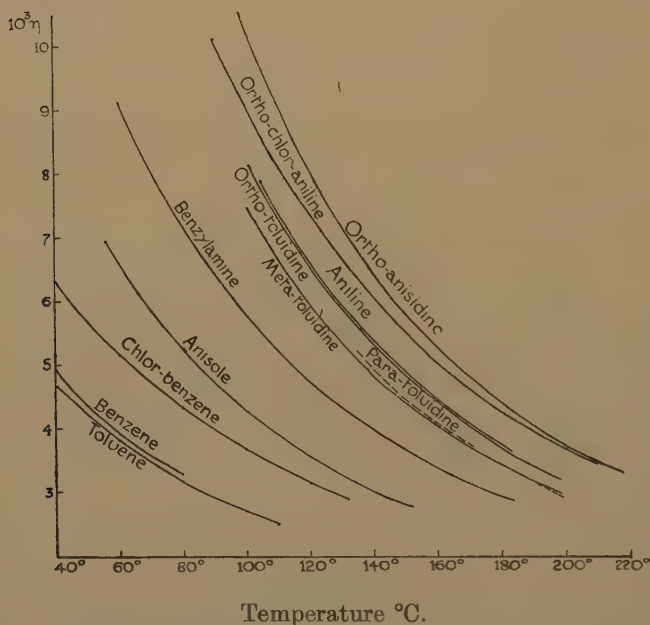
The boiling viscosities and rheochors of the amines are given in Table I. In evaluating these the effect of the benzene ring has been

* Communicated by the Authors.

cancelled out by direct comparison with analogous ring compounds not containing NH_2 . This is desirable because the rheochor of the C_6H_5 -group is found⁽³⁾ to vary somewhat with the electrochemical character of the group to which it is attached. Chlorbenzene⁽⁴⁾ ($R=129.8$) has therefore been selected for comparison with aniline, chloraniline and the toluidines; anisole ($R=141.0$) for anisidine; toluene⁽⁵⁾ ($R=130.2$) for benzylamine (the NH_2 being in the side chain), and cyclohexyl chloride ($R=158.6$) for cyclohexylamine.

No accurate data appear to be available for fatty amines at the boiling point; butylamine and ethylene diamine have therefore been included

Fig. 1.



Viscosity-temperature curves of aromatic amines.

in the study. The agreement with butylamine is close. The value for ethylene diamine is a trifle low, suggesting that the second NH_2 -group is not entirely without influence on the first. It is unlikely that H will have the same rheochor in NH_2 as in CH_2 ; the rheochor of N cannot therefore be evaluated from these data. The ortho effect of the toluidines does not appear in their rheochors. It may be mentioned, however, that data (not yet published) indicate that if the ortho substituent group contains an electron-donating atom, like O in NO_2 , the rheochor is appreciably affected; this suggests a change in constitution^(2a).

TABLE I.

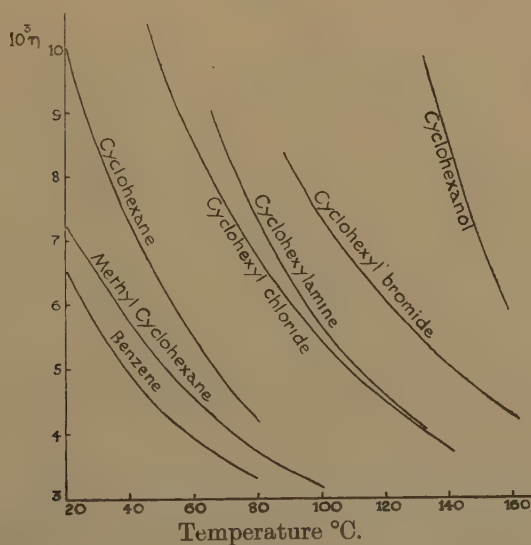
Viscosities and Rheochors of Primary Amines at the Boiling Point.

	$10^3\eta$ at B. pt.	R.	$R(\text{NH}_2)$.
$\text{C}_6\text{H}_5\text{NH}_2$	3.60	123.9	21.4
$\text{C}_6\text{H}_4\text{NH}_2$	3.40	145.1	20.8
$\text{C}_6\text{H}_4\text{CH}_3$ ortho ...	3.16	146.6	20.3
meta	2.96	146.5	20.2
para	2.89	147.3	21.0
$\text{C}_6\text{H}_4\text{OCH}_3$ ortho...	3.28	156.4	20.9
$\text{C}_6\text{H}_{11}\text{NH}_2$	4.00	152.9	21.6
$\text{C}_6\text{H}_5\text{CH}_2\text{NH}_2$	2.83	144.9	20.2
$n\text{C}_4\text{H}_9\text{NH}_2$	2.88	121.1	20.4
$\text{C}_2\text{H}_4(\text{NH}_2)_2$	3.89	87.6	2×20.0
		Mean	20.6

Cyclohexane and Derivatives.

Viscosity-temperature curves are shown in fig. 2. Replacement of hydrogen in cyclohexane by Cl, NH_2 , Br and OH progressively increases the viscosity, whereas CH_3 reduces it. Analogous effects are observed

Fig. 2.



Viscosity-temperature curves of cyclohexane and its derivatives

with benzene (fig. 1). Despite its small mass the hydroxyl group increases the viscosity much more than bromine.

The data in Table II. show that ring formation in the case of mono-substituted derivatives is accompanied by a shrinkage of 5.6 units. Cyclohexane itself gives a contraction of only 4.4 units. A similar anomaly is found with benzene⁽³⁾. Since both cyclohexane and benzene possess zero dipole moments, a connexion between the two phenomena appears possible.

TABLE II.
Viscosities and Rheochors of Cyclohexane and Derivatives
at the Boiling Point.

	$10^3\eta$ at B. pt.	R.	R (Ring).
$C_6H_{11}CH_3$	3.13	160.6	-6.0
$C_6H_{11}Cl$	3.68	158.6	-6.0
$C_6H_{11}Br$	4.24	167.9	-5.2
$C_6H_{11}NH_2$	4.00	152.9	-5.0
$C_6H_{11}OH$	5.88	151.4	-5.9
		Mean	-5.6
C_6H_{12}	4.13	138.4	-4.4

Experimental.

With exception of the cyclohexyl halides, all substances used in this research were obtained from the British Drug Houses and, after drying, were distilled directly into the apparatus.

Cyclohexyl chloride.—Literature refers to the chlorination of cyclohexane⁽⁶⁾ and to the action of fuming hydrochloric acid on cyclohexanol⁽⁷⁾ on the water-bath, for the preparation of cyclohexyl chloride. An attempt to prepare it by the action of phosphorus trichloride on cyclohexanol yielded cyclohexene and a white solid which has not yet been examined. The following method gave a 60 per cent yield. Cyclohexanol was mixed with an equal volume of conc. hydrochloric acid. Anhydrous zinc chloride was added, equal in weight to 25 per cent. of the hexanol. The mixture was heated on the water-bath, a current of hydrogen chloride being passed into the liquid.

After about 30 minutes the liquid suddenly clouded; after a further 60 minutes the passage of hydrogen chloride was stopped and the liquid separated into two layers; the upper one, containing the chloride, was washed with conc. sulphuric acid to remove unchanged cyclohexanol, with sodium carbonate and, finally, with water. On subsequent distillation the middle fraction alone was collected direct in the viscometer and density bottle.

Cyclohexyl bromide.—Re-distilled cyclohexanol was refluxed for two hours with a mixture of conc. sulphuric acid and excess hydrobromic acid, $D=1.49$. The upper layer containing the bromide was separated and dealt with in the manner described above for the chloride.

The results of the density and viscosity determinations are given below, the figures in parentheses referring to data extrapolated at the boiling point. Letters A and B refer to the constants in Sheppard's equation, $\log_{10} 10^3 \eta = \frac{B}{T} + A$.

Aniline.—B. pt. 183° to 183.2° C. at 753.6 mm. $A=-1.1105$; $B=760.3$, from 80° . The equation gives values agreeing well with those of Bramley⁽⁸⁾ at his two highest temperatures, 80° and 125° .

Temp.	105.8	134.2	155.6	169.4	183.2
D	0.9468	0.9217	0.9020	0.8898	(0.8774)
1000 η	7.88	5.708	4.613	4.652	(3.60)

$R=123.9$.

o-Chloraniline.—B. pt. 210.0° C. at 754.6 mm. $A=-0.8963$; $B=690.0$, from 120° .

Temp.	17.25	46.0	60.9	75.1	90.9	120.5
D	1.2117	1.1840	1.1699	1.1560	1.1405	1.1109
1000 η	38.46	19.39	15.14	12.33	10.07	7.195

Temp.	149.4	168.4	189.2	210.0
D	1.0814	1.0616	1.0400	(1.0184)
1000 η	5.432	4.645	3.957	(3.40)

$R=145.1$.

o-Toluidine.—B. pt. 198.7° C. at 749.7 mm. $A=-1.0826$; $B=746.67$, from 100° , at which temperature our extrapolated values for D and η are virtually identical with those of Bingham⁽²⁾.

Temp.	101.8	125.4	137.4	150.0	163.3	175.7	187.8	192.5	198.7
D ...	0.9287	0.9075	0.8968	0.8851	0.8725	0.8608	0.8492	0.8446	(0.8386)
1000 η	8.110	6.189	5.457	4.818	4.249	3.813	3.450	3.320	(3.16)

$R=146.6$.

m-Toluidine.—B. pt. 198.7° C. at 752.1 mm. $A=-1.0560$; $B=720.0$, from 100° .

Temp.	100.7	126.5	139.2	151.3	163.9	175.7	187.0	192.8	198.7
D ...	0.9227	0.9001	0.8885	0.8768	0.8650	0.8539	0.8431	0.8374	(0.8317)
1000 η	7.43	5.58	4.90	4.366	3.904	3.538	3.232	3.088	(2.955)

$R=146.5$.

p-Toluidine.—B. pt. 200° C. at 757.1 mm. $A=-1.1470$; $B=760.7$, over the range studied. Bramley's viscosities at 80° and 100° lie on the

extrapolated curve, but his data depart from linearity above 100°, his observed viscosities being higher than ours.

Temp.	135.6	153.3	167.5	178.0	190.0	200.0
D	0.8890	0.8717	0.8580	0.8477	0.8354	(0.8253)
1000 η	5.18	4.32	3.79	3.466	3.137	(2.89)

$$R=147.3.$$

o-Anisidine.—B. pt. 217.8° C. at 747.8 mm. $A = -1.0922$; $B = 785.2$, from 120°.

Temp.	98.7	124.3	137.7	150.2	163.2	175.5
D	1.0263	1.0013	0.9882	0.9756	0.9627	0.9504
1000 η	10.51	7.67	6.608	5.791	5.104	4.554

Temp.	187.7	200.0	210.0	217.8
D	0.9380	0.9253	0.9150	(0.9072)
1000 η	4.094	3.698	3.413	(3.28)

 $R=156.4.$

Anisole.—B. pt. 151.7° to 152.2° C. at 742.9 mm. $A = -0.9400$; $B = 585.3$, from 50°. Our interpolated value for $10^3\eta$ at 25°, viz. 10.6, is slightly greater than 10.1 found by Baker⁽¹⁰⁾.

Temp.	14.55	55.9	80.5	101.0	114.5	128.0	142.1	152.2
D	1.0035	0.9609	0.9381	0.9182	0.9039	0.8901	0.8745	(0.8635)
1000 η	12.42	6.92	5.236	4.236	3.726	3.320	2.945	(2.735)

 $R=141.0.$

Benzylamine.—B. pt. 183.5° to 184.0° C. at 755.1 mm. $A = -0.9259$; $B = 629.6$, from 60°.

Temp.	21.2	60.9	80.7	101.0	119.0	140.9
D	0.9808	0.9468	0.9297	0.9116	0.8956	0.8758
1000 η	17.8	9.12	7.142	5.721	4.777	3.935

Temp.	156.2	170.0	178.2	184.0
D	0.8620	0.8492	0.8418	(0.8366)
1000 η	3.478	3.122	2.948	(2.83)

$$R=144.9.$$

n-Butylamine.—B. pt. 77.7° C. at 756.4 mm. $A = -1.2765$; $B = 608.5$, over the range studied.

Temp.	18.7	33.0	42.25	51.7	61.7	73.05	77.7
D	0.7479	0.7332	0.7236	0.7136	0.7023	0.6893	(0.6842)
1000 η	6.45	5.13	4.48	3.945	3.475	3.034	(2.88)

$$R=121.1; \text{ calculated } 121.5.$$

Ethylene diamine.—B. pt. 116.7°C . at 758.7 mm. $A = -1.2529$; $B = 718.2$, from 50° .

Temp.	19.6	52.3	63.7	77.5	88.2	99.6	109.3	116.7
D	0.9006	0.8701	0.8597	0.8468	0.8368	0.8260	0.8165	(0.8092)
1000 η	16.86	9.02	7.58	6.246	5.435	4.721	4.223	(3.89)

R=87.6 ; calculated 88.8.

Cyclohexane.—B. pt. 80.7°C . at 759.0 mm. $A = -1.2914$; $B = 675.0$, from 50° . Densities up to 44.6°C . have been recorded by several investigators; our interpolated values are practically identical ^{(11), (12)}.

Temp.	11.2	37.7	54.0	67.0	75.7	80.7
D	0.7859	0.7607	0.7460	0.7337	0.7252	(0.7203)
1000 η	11.58	7.56	5.924	4.926	4.408	(4.13)

$R = 138.4$.

Methyl cyclohexane ⁽¹²⁾.—B. pt. 99.5° to 100.5°C . at 750.5 mm. $A = -0.8470$; $B = 501.3$, from 60° .

Temp.	12.45	50.3	66.8	84.0	97.5	100.5
D	0.7753	0.7425	0.7283	0.7131	0.7013	(0.6987)
1000 η	8.38	5.046	4.254	3.607	3.211	(3.13)

$R = 160.6$; calculated 161.0.

Cyclohexyl chloride.—B. pt. 142.0°C . at 745.7 mm. $A = -0.9253$; $B = 618.66$, from 45° .

Temp.	18.2	47.7	61.1	73.8	86.3	97.8
D	0.9924	0.9644	0.9516	0.9395	0.9275	0.9164
1000 η	16.59	10.19	8.44	7.23	6.258	5.536
Temp.	110.1	121.5	133.4	142.0		
D	0.9013	0.8932	0.8815	(0.8730)		
1000 η	4.891	4.391	3.966	(3.68)		

$R = 158.6$; calculated 159.0.

Cyclohexyl bromide.—B. pt. 161.5° to 162.0°C . at 733.5 mm. $A = -0.8222$; $B = 630.4$, from 90° .

Temp.	16.2	88.6	106.5	121.0	134.4	155.0	162.0
D	1.3363	1.2462	1.2239	1.2069	1.1901	1.1637	(1.1546)
1000 η	26.0	8.35	6.869	5.975	5.30	4.46	(4.24)

$R = 167.9$; calculated 167.5.

Cyclohexylamine.—B. pt. 132.0° to 132.5°C . at 745.0 mm. $A = -1.1915$; $B = 727.3$, from 60° .

Temp.	15.1	43.5	66.4	82.1	101.8	113.7	125.8	132.5
D	0.8722	0.8467	0.8255	0.8117	0.7934	0.7826	0.7718	(0.7657)
1000 η	24.72	13.07	8.935	7.117	5.543	4.853	4.269	(4.00)

$R = 152.9$; calculated 152.3.

Cyclohexanol.—B. pt. 158.5°C . at 757.9 mm. $A = -2.6947$; $B = 1494.3$ from 100° .

Temp.	99.4	116.5	124.7	132.0	139.3	147.4	151.0	158.5
D	0.8817	0.8646	0.8565	0.8489	0.8412	0.8325	0.8284	(0.8203)
1000 η	20.77	13.73	11.52	9.88	8.51	7.25	6.756	(5.89)

$R = 151.4$; calculated 151.7.

References.

- (1) Friend, 'Nature,' cl. p. 432 (1942); Friend and Hargreaves, Phil. Mag. xxxiv. p. 643 (1943).
- (2) Using data of Thorpe and Rodger, Phil. Trans. clxxxv. p. 571 (1895), for benzene and toluene.
- (2 a) As suggested by Sidgwick and Callow, J. Chem. Soc. cxxv. p. 527 (1924); Sidgwick, "The Electronic Theory of Valency," p. 147 (1927).
- (3) Unpublished data.
- (4) Calculated from Titani's data for viscosity (Bull. Chem. Soc., Japan, ii. p. 95 (1927)) and Young's densities (given by Landolt-Börnstein).
- (5) Calculated from data given by Thorpe and Rodger, ref. 2.
- (6) Fortey, J. Chem. Soc. lxxiii. p. 940 (1898).
- (7) Markownikow, *Annalen*, cccii. p. 11 (1898).
- (8) Bramley, J. Chem. Soc. cix. p. 20 (1916).
- (9) Bingham, van Klooster and Kleinspehn, J. Phys. Chem. xxiv. p. 1 (1920).
- (10) Baker, *ibid.* ci. p. 1049 (1912).
- (11) Eg. Vogel, J. Chem. Soc. cli. p. 1323 (1938).
- (12) The viscosities are 1 per cent. higher than those quoted by Evans, II Congr. mondial pétrole, ii. sec. 2, p. 933 (1938).

The Technical College,
Birmingham.

VIII. Modifications of Appleton's Method of Measuring the Mutual Conductance of a Valve.

By L. F. BATES, Professor of Physics, and W. F. LOVERING,
University College, Nottingham*.

[Received January 4, 1944.]

MANY years ago Sir Edward Appleton designed the extremely neat method for finding the mutual conductance, g_m , of a triode, for which the circuit is shown in fig. 1. It can readily be shown that if the valve is biased to work on the straight portion of its characteristic, so that $I_a = k(V_B + \mu V_g)$, the resistance R may be so chosen that there is no change in the reading of the milliammeter, M , on opening or closing the switch K . In these circumstances and on the assumption that V_B possesses negligible resistance, $g_m = k\mu = 1/R$; for, the increase in grid potential is equal to Ri , where i is the current flowing through R when K is closed, and this gives rise to an increase i in the plate current such that $k\mu Ri = i$, so that there is no change in the milliammeter deflection.

* Communicated by the Authors.

As a laboratory experiment the milliammeter may profitably be replaced by a moving magnet galvanometer of the Kelvin type with a control magnet, so that the latter may be arranged to give a temporary zero in any convenient position with a lamp and scale. The main errors in the measurement arise from the internal resistance of the battery and from slow changes in filament activity of oxide-coated cathodes, since this activity is increased with increase in plate current.

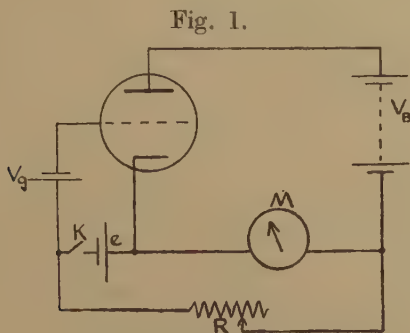
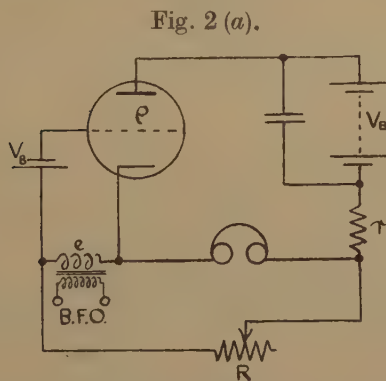


Diagram showing the original Appleton circuit.



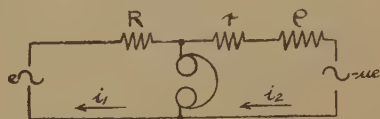
Modified Appleton circuit using beat frequency oscillator and telephones.

We have found that the indicating instrument may be replaced with advantage by telephones, using a beat frequency oscillator, of the type now supplied to university institutions, in place of the battery e and switch K . It is desirable, however, to connect the B.F.O., as shown in fig. 2 (a), in the circuit through a matching transformer to avoid distortion which may occur when the B.F.O. output valve is not operating under satisfactory conditions, and to prevent the steady plate current of the valve under test from damaging the B.F.O. itself. The telephone

impedances cannot interfere with the action of the circuit because no alternating current flows through them when $g_m = 1/R$, assuming $r = 0$.

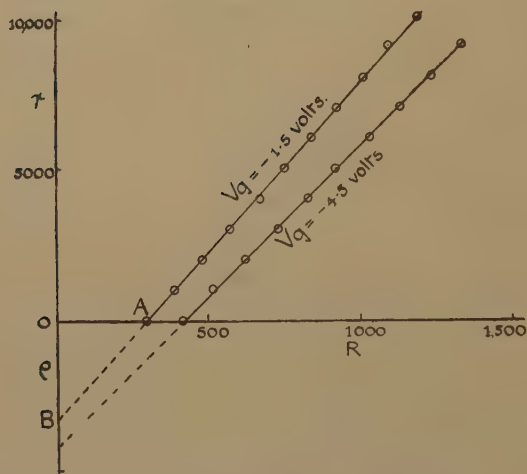
This arrangement enables us to short-circuit any resistance in V_B by a condenser of large capacity. It also enables us to provide an instructive laboratory exercise if we insert a variable resistance r in the plate

Fig. 2 (b).



Equivalent circuit of fig. 2 (a).

Fig. 2 (c).



Graph showing relation between r and R of fig. 2 (a) for chosen values of V_B and V_G with a M.L.4 valve, $V_B = 100$ volts. $OA = 1/g_m$; $OB = \rho$; slope $BA = \mu$.

circuit. For, referring to fig. 2 (b), we then have at balance, $i_1 = i_2$, so that

$$\frac{e}{R} = \frac{\mu e}{\rho + r}, \quad \text{or} \quad r = \mu R - \rho,$$

whence

$$g_m = \frac{1}{R} + \frac{r}{\rho R}$$

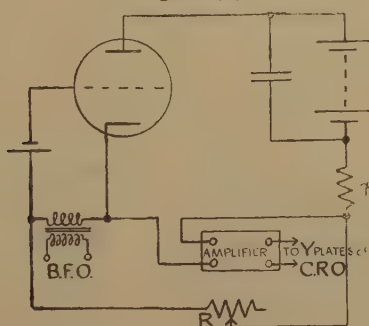
If therefore, we make $r = 1000, 2000$, etc., ohms and find the corresponding values of R , we may plot r against R , as in fig. 2 (c), and so read off the values of μ , g_m and ρ from the graph. Using a M.L.4 valve with $V_B = 100$ volts and $V_G = -1.5$ volts, the values obtained for these

quantities were, 10.5, 3.48 milliamperes per volt and 3120 ohms respectively.

It is essential to use weak signals from the B.F.O., *e. g.* less than 0.5 volt, when R may be found with an accuracy of 1 part in 200. If strong signals are used, a second harmonic caused by curvature distortion makes detection of the balance somewhat uncertain, so that it is better, if greater accuracy is required, to provide the telephones with an amplifier.

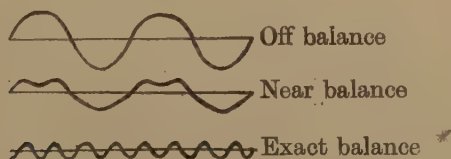
In spite of its great simplicity, the Appleton method for g_m appears to be very much less well known than it should be, and the following experiments were devised as a result of arranging the method in the

Fig. 3 (a).



Circuit for using C.R.O. in place of telephones, etc.

Fig. 3 (b).



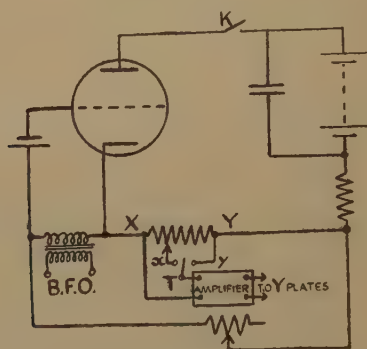
C.R.O. traces obtained with circuit of fig. 3 (a).

form of a lecture demonstration with a cathode ray oscillograph as indicating instrument. There are several ways in which this can be done. The most obvious method is to replace the telephones of fig. 2 (a) with an amplifier and C.R.O., preferably a Graphoscope or a C.R.O. with its own amplifier incorporated, as in fig. 3 (a). The X plates are connected to a suitable time base and the Y plates record an alternating potential when balance is not obtained.

Now, with the last circuit, only the fundamental frequency of the B.F.O. can be balanced out by adjusting R . The second harmonic injected into the C.R.O. by the valve cannot be so eliminated, and we therefore get a series of traces similar to those depicted in fig. 3 (b). The experiment thus provides a beautiful method of illustrating the

presence of a second harmonic. Moreover, it enables us to measure the percentage distortion if we change the circuit to that of fig. 3 (c), where the resistance XY acts as a potential divider. If the tongue T is joined to y , the circuit can be balanced by altering R as before. The amplitude of the C.R.O. trace must be noted, and the tongue T is then joined to x and the H.T. circuit is broken at K , when the potential divider can be adjusted to provide a trace of the same amplitude. With the y connection and K open the whole voltage or amplitude of the second harmonic is measured; with the x connection that fraction of the amplitude of the fundamental equal to that of the full second harmonic is measured. The H.T. connection must be broken in order to leave only the fundamental B.F.O. component. The ratio of the resistance of Xx to that of XY gives the ratio of the amplitudes of the second harmonic and fundamental.

Fig. 3 (c).



Circuit for measuring percentage distortion.

We now come to an arrangement which was, in fact, the first extension of the Appleton circuit of fig. 1 which we devised. In fig. 4 (a), two equal resistances r and r^1 are inserted in the plate and subsidiary circuits. These are connected to a C.R.O. with amplifiers as shown. When the switch K is open the signal from the B.F.O. passes through r and r^1 in series, and as there can be no phase difference between the potential drops across these resistances, a linear trace is provided by the C.R.O. The trace will make an angle of 45 degrees with the vertical if the X and Y amplifications are equal, but this is not essential, as it is only necessary to record the direction of the trace by suitable means.

If K is now closed we have the necessary connections for the determination of g_m . We have merely to vary R_1 until equal alternating potential differences exist between the ends of r and r^1 , shown by the formation of a linear trace *at the same angle* as before, when, clearly, no alternating current can pass through K . Actually, this does not constitute a true Appleton method of measurement, because r may introduce

appreciable resistance into the plate circuit, and we have the potential difference across the Y plates equal to the potential difference across the X plates, so that

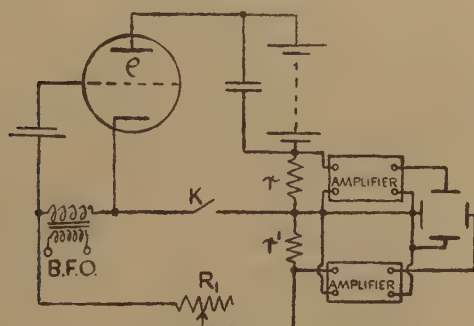
$$\frac{\mu er}{\rho + r} = \frac{er}{R_1 + r},$$

whence
$$\frac{\mu}{\rho + r} = \frac{1}{R_1 + r}, \quad \text{or} \quad \frac{\mu}{\rho} = \frac{\rho + r}{\rho} \frac{1}{R_1 + r},$$

because $r = r'$. Hence,
$$g_m = \left(\frac{1}{R_1 + r} \right) \left(1 + \frac{r}{\rho} \right).$$

With a M.L.4 valve $R_1 + r$ can be found to an accuracy of 1 part in 20 only, although the experiment nevertheless makes a most effective lecture demonstration. The amplifiers on the X and Y plates should be

Fig. 4 (a).



Circuit for measurement of g_m using linear trace.

similar, as we obtain elliptical traces if any phase shifts are introduced. It should be noted that large signals are necessary or the amplification must be high, that the value of g_m so obtained is an average over a considerable stretch of a characteristic, and μ and ρ are likewise average values. Owing to curvature of the characteristic with strong signals, the outer portion of the trace obtained may be non-linear, but there is usually enough linear trace for reasonable accuracy in setting. A convenient B.F.O. frequency is about 2000 cycles per sec. for ordinary working.

The circuit of fig. 4 (b) shows how the last circuit may be adapted for the measurement of ρ . It is convenient to insert a reversing switch S in conjunction with either the X or the Y plates if the circuits of figs. 4 (a) and 4 (b) are combined with a three-way key which enables the B.F.O. to be inserted either to the left or to the right of the filament terminal A, since the potential differences between the ends of r and r^1 are 180 degrees out of phase in the circuit for ρ , and it is desirable to

work with the 45 degree trace always in the same direction. With the full line connections in fig. 4 (b) we have, for balance, $R_1 = \rho$, and, with the addition of the connection represented by the broken line, $R_1 = \rho/1 + \mu$, so that ρ and μ may be found, if necessary, to an accuracy of about 1 part in 20.

Now, the last method requires the equality, or, at any rate, the fixing of resistances r and r^1 . This may be avoided by using the circuit of fig. 5 (a) in which the switch K of fig. 4 (a) is replaced by a simple form

Fig. 4 (b).

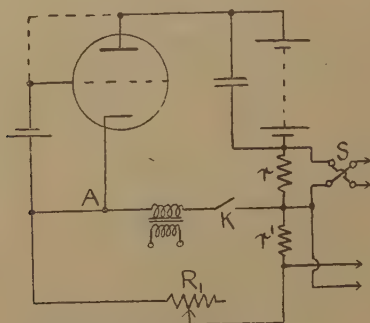
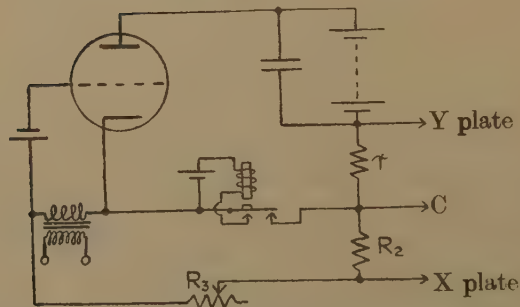
Circuit for measurement of ρ .

Fig. 5(a).

Circuit for the double trace method of measurement of g_m .

of contact breaker, merely a second contact on a buzzer operating a few hundred times per sec. When the contact is open, whatever the relative values of r and R_2 , the C.R.O. gives a linear trace whose slope is determined only by the potential differences across r and R_2 . When the contact is closed a second line is recorded. In general the second line is separate from the first and of different slope, but by adjustment of R_2 balance may be obtained, and when this is the case the two lines must have the same slope. The appearance and the separation of the lines depend on the operating conditions, and they are seen simultaneously

because of persistence of vision. Should the valve be driven beyond "cut-off" by a strong signal during each cycle, rectification produces a sideways displacement of one of the traces exactly as is also obtained by changing the grid bias, etc. This is a distinct advantage as the balance may then be more easily decided. Since r and R_2 can be allowed to have practically any values we please, the former can be made small, *e. g.* of the order of 20 to 50 ohms, and so be negligible compared with

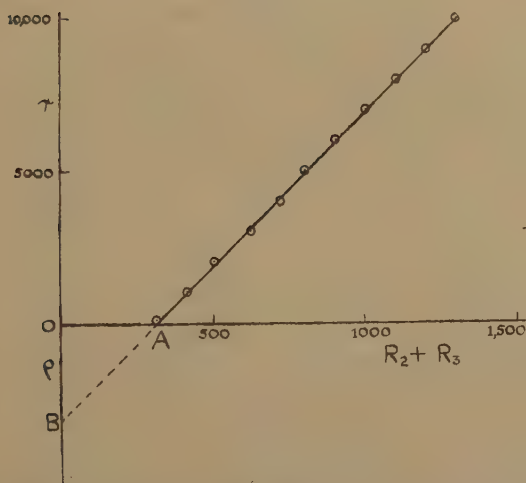
Fig. 5 (b).



R too small. R correct. R too large.

Diagrams of double trace records.

Fig. 6.



Results obtained by the double trace method of fig. 5 (a) for a M.L.4 valve, $V_B=100$ volts; $V_G=-4.5$ volts. $OA=1/g_m$; $OB=\rho$; slope of $BA=\mu$.

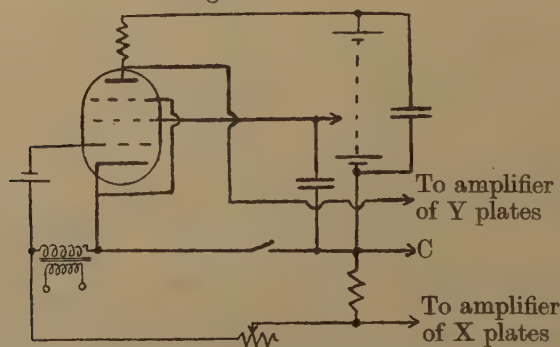
ρ . We may then write with little error $g_m = \frac{1}{R_2 + R_3}$. The error is, in fact, $\frac{r}{\rho} \cdot g_m$, which is of the order of 1 to 2 per cent. of g_m .

The main advantage of the method is that the amplifier may be run at high gain to give long lines on the C.R.O. screen. We found it easy to obtain clear traces 6 to 7 cm. long and about 1 to 2 mm. apart, when the value of R_3 could be found to within 5 ohms, *i. e.* to about 1 part in 80 for $R_2 + R_3$. The traces usually have the forms shown in fig. 5 (b). It is, of course, essential to work with a B.F.O. frequency which obviates phase shifts when amplifiers are used.

A typical set of results with a M.L.4 valve, obtained by giving r stated values as in an earlier portion of this work, are shown graphically in fig. 6. The values of g_m , ρ and μ read from the graphs are 3.28 milliamperes per volt, 3050 ohms and 10.0 respectively for 100 volts H.T. and -4.5 volts grid bias. The whole makes a very effective laboratory or lecture experiment.

So far we have confined our attention to experiments with triodes, but the methods may be used for screen grid valves and pentodes. The circuit for a pentode is shown in fig. 7. Owing to curvature of the characteristic the trace obtained is non-linear and the experimenter usually measures g_m corresponding to zero or positive grid voltages, *i. e.* in the region where the valve characteristic is practically linear. With a M.S.P.4 valve, anode potential 250 volts, screen grid 100 volts, we

Fig. 7.



Circuit for mutual conductance of a pentode.

found g_{ac} to be 5.2 milliamperes per volt, with an accuracy of 1 part in 20.

SUMMARY.

Several new modifications of Appleton's method for measuring the mutual conductance of a valve are described. Most of them require the provision of a cathode ray oscillograph and are suitable for lecture demonstrations. The best so far devised gives a two line trace on the C.R.O. screen and enables accurate measurements of ρ , μ and g_m to be made.

Reference.

Appleton, E., Proc. Camb. Phil. Soc. xx. p. 239 (1920-21).

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

IX. *Molecular Weight, Gram Molecule and Avogadro's Number.*

By W. E. DUNCANSON, M.Sc., Ph.D.*

[Received April 27, 1943.]

1. *Introduction.*

THE three allied physical quantities known as the molecular weight, the gram molecule and Avogadro's number are so closely related that at times there is confusion as to which should be used in a specific instance. We shall here attempt to preserve a sharp distinction between these three quantities and their use in physical theory.

The nomenclature is, in some ways, unfortunate but will be dealt with here only in so far as it affects the immediate discussion. A more detailed discussion of terminology is given in the following paper.

The symbols A , \mathcal{M} , \mathcal{N} and m will be used throughout to denote respectively, the molecular weight, the gram molecule, Avogadro's number and the mass of an individual molecule. \mathcal{M} is preferred for gram molecule rather than for the molecular weight, as \mathcal{M} suggests a mass whereas we shall show below that molecular weight is a ratio. A has been chosen for this ratio as it is merely a slightly more generalized use of the symbol which normally represents atomic weight.

2. *Definitions.*

Before an adequate discussion can be given the quantities under consideration must be defined.

Molecular Weight.—There are two general ways in which molecular weight may be defined. These may be called the macroscopic or bulk method, and the microscopic or molecular method. The microscopic definition is the better as it is the more fundamental, while the macroscopic method is limited in that it requires the substance to be in the gaseous form and assumes an ideal gas.

The macroscopic definition does not necessarily refer to atoms or molecules, but its historical background assumes their existence. It may be stated thus :

The molecular weight of a substance is 32 times the ratio of the mass of that substance to the mass of O_2 which, in gaseous form, will occupy the same volume under the same conditions of temperature and pressure.

* Communicated by the Author.

where m is the mass of one molecule of the substance. From these two equations we see that \mathcal{N} is a dimensionless number, and \mathcal{M} a quantity with the dimensions of mass.

Unfortunately, there is no term of quantity for this particular number of molecules denoted by Avogadro's number. We can refer to numbers such as a thousand, a million, a lac (generally of rupees !), but an Avogadro's number of molecules is not the most pleasing of terms to use. It would be convenient if the term *mole of molecules* could be used for this purpose, but mole is used synonymously for gram molecule. The lack of such a term to denote Avogadro's number is unfortunate as it leads to the use of the term gram molecule irrespective of whether is meant a particular number of molecules or the mass of these molecules.

We suggest that the word "*Avogadron*" be used as the term of quantity denoting a number equal in magnitude to Avogadro's number. We shall now refer to an "*avogadron of molecules*."

3. Discussion.

This lack of a term to express Avogadro's number has been one of the reasons, in the past, for the failure at times to distinguish it from the gram molecule. But there is a further reason. A particular number of molecules of a substance can be most conveniently obtained by weighing out a certain calculable quantity of the substance. For example :

If \mathcal{M}_x and m_x are the gram molecule and the mass of a molecule of the substance, and \mathcal{M}_0 and m_0 the corresponding quantities for atomic oxygen, then by definition

$$\mathcal{N} = \frac{\mathcal{M}_x}{m_x} = \frac{\mathcal{M}_0}{m_0}.$$

Hence the mass \mathcal{M}_x of \mathcal{N} such molecules is given by

$$\begin{aligned}\mathcal{M}_x &= \mathcal{M}_0 \times \frac{m_x}{m_0}, \\ &= \mathcal{M}_0 \times \frac{A_x}{16},\end{aligned}$$

where A_x is the molecular weight of the substance, and \mathcal{M}_0 by definition equals 16 gm. This then gives us a means of calculating the mass of \mathcal{N} molecules of the substance, and therefore by the simple method of weighing \mathcal{N} molecules may be selected.

But this method of selecting \mathcal{N} molecules by weighing does not mean that a quantity of the dimensions of mass always appears in our calculations. A particular number of molecules can be selected by weighing because of previous determinations; and, be it noted, this particular number can be chosen, to the degree of accuracy permitted by weighing, without knowing the actual magnitude of this number. An example or two will illustrate this point more clearly.

Quite often it is most convenient, and may be more accurate with the apparatus available, to select a definite volume of a substance by

weighing, but it is obviously wrong to say that because this volume has been chosen by weighing that a quantity of the dimensions of [mass] has been introduced into the work. The actual mass to be weighed out is known from a previous determination of the density. Hence the complete set of processes involved in "selecting a volume by weighing" are the measurement of a mass, the measurement of a density, and the mathematical operation of dividing the result of the mass measurement by that of the density measurement.

Again, the uniform velocities of bodies between two points can be measured by a stop watch only—assuming that the distance has been measured in some previous experiment and is known!

In other words, we must not deceive ourselves into thinking we are making certain measurements only when the knowledge of other quantities is assumed, such as the density in our first example and a length in our second. Similarly, in selecting an avogadron of molecules by weighing we assume previous knowledge of the gram molecule for oxygen and the molecular weight of the particular substance in question.

Thus, just because there has only been one term to denote both a particular number of molecules and their mass, and just because both these quantities are generally selected by weighing, it must not be assumed that it is the gram molecule (that is, a quantity whose dimensions are [M]) which is always involved in the theory.

Furthermore, as the gram molecule and the molecular weight have the same numerical value they are often used in formulæ without due care being taken as to which one is actually involved.

This lack of precision in the use of molecular weight, gram molecule, and Avogadro's number is very much to be deplored, not from a purist point of view, but because of the confusion and error to which it leads even in many of the standard text-books!

In the light of these remarks let us examine some of the thermodynamical quantities which involve one or other of these quantities.

4. *Temperature* *.

From simple kinetic theory considerations we have

$$pV = \frac{1}{3}nm\overline{c^2}, \quad (3)$$

where p and V are respectively the pressure and volume of n molecules of a gas each with a mass m and a mean square velocity $\overline{c^2}$.

If an avogadron of molecules is chosen then n is replaced by \mathcal{N} and V by \mathcal{V} to emphasize that a particular volume—namely the volume of \mathcal{N} molecules—is under consideration. Equation (3) then becomes

$$p\mathcal{V} = \frac{1}{3}\mathcal{N}m\overline{c^2}. \quad (4)$$

* This has been dealt with by the author in less detail and chiefly from the point of view of dimensions in previous papers (Proc. Phys. Soc. liii. p. 432 (1941) and liv. p. 504 (1942).

The method of introducing temperature into this equation is controversial. Two possibilities present themselves:—

(a) The mean square of the velocity $\overline{c^2}$ is identified with the temperature which we shall denote by θ' , omitting for the time being any scale factor to convert the unit of temperature to a more convenient size*.

Equation (4) is then written

$$p\mathcal{V} = \frac{1}{3}\mathcal{N}m\theta'. \quad (5)$$

Replacing $\mathcal{N}m$ by \mathcal{M} from equation 2, we get

$$p\mathcal{V} = \frac{1}{3}\mathcal{M}\theta'. \quad (6)$$

This is consistent with the statement that $\frac{p\mathcal{V}}{\mathcal{M}}$ is a measure of temperature.

But this method of approach does not lead to the usual concept of temperature. In a mixture of two different gases in equilibrium, according to the equipartition of energy, the lighter gas will have the greater velocity and hence the higher temperature.

(b) The alternative is to identify the mean energy $\frac{1}{2}m\overline{c^2}$ with the temperature which we shall denote by θ . Then equation (4) becomes

$$p\mathcal{V} = \frac{2}{3}\mathcal{N}\theta, \quad (7)$$

again ignoring any scale factor to convert θ to appropriate units. θ is now equal to the mean energy of the molecules; therefore, because of the equipartition of energy, we now have two different gases reaching

the same temperature when mixed. In this case $\frac{p\mathcal{V}}{\mathcal{N}}$ is the measure of temperature; that is the first method gives $\frac{p\mathcal{V}}{\mathcal{M}}$ and the second $\frac{p\mathcal{V}}{\mathcal{N}}$

as the measure of temperature. Now the second method is the usual one, but owing to the confusion in the use of gram molecule, and hence the use of \mathcal{M} instead of \mathcal{N} , temperature is often said to be measured in terms of energy per gram molecule, thus making the erroneous deduction

that the dimensions of temperature are $\left[\frac{\text{energy}}{\text{mass}} \right]$.

But the θ we have introduced is not expressed in the usual units, so let us introduce a scale factor to convert θ to more suitable units. This is achieved by putting

$$\theta = \frac{3}{2}kT. \quad (8)$$

The temperature scale denoted by T is called the absolute scale. Equation (7) now becomes

$$p\mathcal{V} = \mathcal{N}kT = \mathcal{R}T, \quad (9)$$

where

$$\mathcal{R} = \mathcal{N}k. \quad (10)$$

* Cf. a discussion of scale factors in a previous paper (Proc. Phys. Soc. liii. p. 432 (1941)).

Frequently this equation is written in the forms

$$pV=RT \quad \text{or} \quad pv=RT,$$

where V (or v) is called the specific volume, or the volume per unit mass. But it should not be denoted by the same symbol as that employed for volume. It is not a volume, any more than density is a mass. As defined, specific volume is a volume divided by a mass (the reciprocal of density) and should be given a distinct symbol*, for example \mathcal{V} , then equation (16) should be written as it stands or as

$$p\mathcal{V}=RT, \quad (17)$$

$$\text{or, of course,} \quad \frac{p}{\rho}=RT, \quad (18)$$

where ρ is the density.

6. Molar Heats.

When we come to consider the definition of molar heat (or molecular heat) we again meet with looseness in terminology. Some define molar heat as the product of the specific heat and the gram molecule, while others define it as the product of the specific heat and the molecular weight. In the first case molar heat has the dimensions of [specific heat \times mass], whereas in the second case the dimensions are the same as those of specific heat because molecular weight is a ratio. Let us consider which is the appropriate definition.

Let us take a mass M of a gas which has a volume V_1 at a pressure p and temperature T_1 . When the temperature is raised to T_2 at constant pressure let the new volume be V_2 . If q_p is the heat required to raise the temperature from T_1 to T_2 at constant pressure, and q_v the heat required for the same process at constant volume, then

$$\begin{aligned} q_p &= q_v + p(V_2 - V_1) \\ &= q_v + MR(T_2 - T_1). \end{aligned} \quad (19)$$

From the definition of the specific heats we have

$$c_p = \frac{q_p}{M(T_2 - T_1)} \quad \text{and} \quad c_v = \frac{q_v}{M(T_2 - T_1)},$$

and therefore equation (19) can be written

$$c_p - c_v = R = \frac{\mathcal{R}}{\mathcal{M}}. \quad (20)$$

$$\text{If we put} \quad C_p = c_p \times \mathcal{M} \quad \text{and} \quad C_v = c_v \times \mathcal{M}, \quad (21)$$

$$\text{then we have} \quad C_p - C_v = \mathcal{R}. \quad (22)$$

This usual relation for the difference between the molar heats has been obtained by defining them, in equation (21), as the *product of the appropriate specific heat and the gram molecule*—not the molecular weight. Therefore, the molar heat has the dimensions of [specific heat \times mass]. If it is agreed that temperature has the dimensions of [energy] then specified heat has the dimensions $[M^{-1}]$, and hence the molar heat is dimensionless.

* See following paper.

Let us consider the units in which the molar heat is to be expressed. It is defined as the product of a specific heat and a mass, consequently it is in the nature of a thermal capacity. Now the thermal capacity of a body is expressed in cal. per degree (or other similar units). For example, the thermal capacity of a particular calorimeter is found to be x cal. per $^{\circ}\text{C}.$; or the thermal capacity of 24 gm. of a particular substance is found to be y cal. per $^{\circ}\text{C}.$ In other words, mass does not enter into the units of thermal capacity, except in so far as it is involved in the calorie.

When we turn to molar heats we find a similar state of affairs. The molar heat of O_2 is just the thermal capacity of 32 gm. of O_2 ; the molar heat of Xe is the thermal capacity of 131 gm. (approx.) of Xe, and so on. Therefore, the appropriate units for molar heat are cal. per degree. But, it may be objected, a new unit of mass * is introduced in each case, equal to 32 gm. in the case of O_2 , and equal to 131 gm. in the case of Xe, and the conclusion is then drawn that the molar heat is a specific heat and not a thermal capacity.

The consequences of adopting such a new unit of mass are :

- (a) It means a different unit of mass for each substance.
- (b) The new unit of mass is not used consistently throughout, because the calorie (or, more fundamentally, the erg) is defined in terms of a mass in *grams*, not in this new unit. This use of mixed units, especially when the magnitude of one of the units varies from substance to substance, is not conducive to precise statement.

To conclude, we emphasize again that the molar heat is the product of the specific heat and a mass (expressed in gm., be it noted), and therefore the units of the molar heat of a particular body or collection of bodies under consideration are just cal. per degree.

7. *Latent Heats.*

The treatment of latent heat follows closely similar lines to that of molar heat and need not be considered in detail.

If l is the quantity of heat required to change the state of unit mass of a substance without change of temperature, then in the c.g.s. system the units of l are cal. per gm. The molar latent heat, L , is defined as the product of the latent heat and the gram molecule, \mathcal{M} ; that is

$$L = \mathcal{M} \times l. \quad \dots \dots \dots (23)$$

Thus defined the molar latent heat is expressed in cal.

University College, London,
c/o University College of North Wales,
Bangor.

* This new unit is called by many writers the gram molecule—and it should be noted that the term in this sense is being used to denote a *unit* and not a physical quantity. A further discussion of this point is given in the following paper.

X. *Terminology and Symbols in Physics.*

By W. E. DUNCANSON, M.Sc., Ph.D.*

[Received April 27, 1943.]

TERMINOLOGY.

1. *Introduction.*

In the preceding paper there has been discussed the confusion in the use of the quantities molecular weight and gram molecule on the one hand, and between gram molecule and Avogadro's number on the other. The first arises because the two quantities have the same numerical value and, as the term molecular weight suggests a weight (or mass) instead of a ratio, unless care is taken it is apt to be used when in fact gram molecule is intended. Confusion occurs in the second case because there is no term of quantity to denote Avogadro's number, and as a result the term gram molecule is used loosely to indicate either a particular number of molecules or their mass.

Moreover, the term gram molecule is often used in two entirely different senses. At times it is used to denote a physical quantity, analogous to the use of the terms density or velocity; at other times it is adopted as a unit of mass and reference is made to an amount such as 4 gram molecules. This is analogous to saying a quantity has a value of 4 densities! Fortunately, the word *mole* is frequently chosen as the term to represent this unit of mass, and it is most desirable that if such a unit is to be used at all a separate term such as *mole* should be set apart for it and used for that purpose only.

The difficulties and errors resulting from loose or wrong terminology are not confined to these two examples and we shall review some of the more common quantities where this problem arises.

Let us first consider the conditions that must be satisfied in order that a consistent and self-evident scheme of nomenclature may be built up.

(1) The first step is that, having unambiguously defined a quantity, the name given to it should bear some relation to the definition, or at least, should not be in direct contradiction to the definition, as in the case when a quantity, defined as a ratio of two masses (or weights) is called a *molecular weight*.

2. When introducing a term it must be made quite clear to which particular physical quantity it refers. It should in no case be applied to two or more different quantities, as when the term gram molecule is used to refer both to a particular number of molecules and also to their

* Communicated by the Author.

mass; nor should the same term refer to both a physical quantity and the unit in which it is measured, as we again have in the case of gram molecule.

3. When qualifying words or phrases such as *relative* or *specific* are used a definite and consistent meaning should be applied to them so that this meaning is immediately obvious.

In the following the nomenclature of a number of quantities will be discussed and suggestions will be made as to modifications or alterations in the nomenclature to be consistent with the definitions, but with as little deviation as possible from common usage. For the sake of clarity the definitions will be kept as simple as possible. Each section will be headed with the name now in common use.

2. Density.

In this case the nomenclature is fairly satisfactory and is treated here to illustrate the general principle on which the remainder of the paper is based.

Density is defined as the mass of a given substance per unit volume.

Often in physics it is desired to refer the value of a quantity for a particular substance to the value of that quantity for some reference substance. In the case of density the reference substance generally chosen is water and we define

Relative density of a substance as the ratio of the mass of that substance to the mass of an equal volume of water.

In the case of gases it is sometimes more convenient to adopt hydrogen as the reference substance. If necessary the term *relative gas density* could be used for this purpose, but generally there is no doubt as to whether water or hydrogen is chosen as the reference substance.

The word *relative* is used throughout this paper to denote the ratio of the value of a quantity for a particular substance to the value of the same quantity for some reference substance, although later in the paper *specific* is considered as a possible alternative.

For the sake of distinction we shall refer to quantities, such as density, as the fundamental quantities, and the ratios, such as relative density, we shall call derived quantities.

3. Specific Volume.

This term is used to denote the volume per unit mass, that is, the reciprocal of density. The use of the word *specific* has no precise meaning in physics. In some cases, such as specific volume and specific charge, it refers to the value of the particular quantity *per unit mass*; but this is by no means the general use as can be seen in the case of specific resistance and specific inductive capacity. With such an indefinite use of the word *specific* the term specific volume is too liable to suggest a

quantity of the dimensions of [volume] rather than of $\left[\frac{\text{volume}}{\text{mass}} \right]$. It

would be much more satisfactory to have a separate name for this quantity to distinguish it from an ordinary volume. As it is the reciprocal of a density the terms *rarity* or *sparsity* suggest themselves. We choose *rarity* and define the

Rarity of a substance as its volume per unit mass.

4. *Specific Heat.*

This has already been discussed by the author ⁽¹⁾ from the point of view of confusion of the fundamental and the derived quantities consequent on the use of only one name for both quantities. There the emphasis was not on terminology and the unwieldy term thermal capacity per unit mass was used for the fundamental quantity. Its use was justified in that respect as its meaning was quite definite and it was desired not to confuse the point under discussion at the time by introducing an entirely new term. However, thermal capacity per unit mass is not a particularly convenient phrase for a quantity in common use. In a discussion of possible names for such a quantity Mr. D. O. Wood suggested the word *thermity* as one which might be acceptable for the fundamental quantity. It has the same form as its prototype *density* and is also compact.

Thermity of a substance is then defined as the quantity of heat required to raise unit mass of the substance through unit temperature ; and

Relative thermity of a substance is the ratio of the quantity of heat required to raise a given mass of the substance through a given temperature to the quantity required to raise the same mass of water through the same temperature.

5. *Dielectric Constant and Permeability.*

Guggenheim ⁽²⁾ and the author ⁽¹⁾ have discussed the use of the fundamental and derived quantity in this case. Guggenheim has suggested the term *permittivity* for the fundamental quantity and *specific inductive capacity* or *dielectric constant* for the derived quantity. In conformity with the other suggestions in this paper it would be more uniform to adopt *permittivity* and *relative permittivity*.

The *permittivity* (k) of a substance would then be defined by the equation

$$k = \frac{Q_1 Q_2}{F r^2},$$

where F is the force between two charges Q_1 and Q_2 placed at a distance r apart in the substance. Or, the permittivity of a substance is the square of the value of two similar and equal charges which, when placed at unit distance apart in the substance, repel one another with unit force ; and

The *relative permittivity* is the ratio of the permittivity of the substance to that of a vacuum.

Similarly in the magnetic case we have *permeability* and *relative permeability*.

6. Avogadro's Number.

Avogadro's number is a particular number chosen so that this number of O_2 molecules has a mass of 32 gm. A name is needed to represent this particular number, and we have suggested in the preceding paper the word *avogadron* as the term of quantity, so that reference may be made to an avogadron of molecules, by which is meant a number equal to Avogadro's number.

7. Gram Molecule.

With present usage gram molecule may convey three distinct meanings. It may denote

- (a) a certain number of molecules,
- (b) the mass of this particular number of molecules, and
- (c) a unit of mass.

To resolve this confusion we suggest

(i.) as already put forward in the previous section, that *avogadron* be adopted to denote the particular number in question,

(ii.) that *molar mass** be used to denote the mass of an avogadron of molecules,

(iii.) that mole be retained for the new unit of mass—that is, the mass of an avogadron of molecules—although, as pointed out in the preceding paper, to employ this unit at all is not very desirable.

The advantage of these suggestions is that molar weight (if not molar mass) and mole are already in general use, and hence there is no need for the term gram molecule at all especially as it does not make clear precisely for which purpose it is being used.

8. Molecular Weight.

Two definitions of this quantity have been given in the previous paper, the microscopic definition in terms of the ratio of the mass of the molecule to the mass of an oxygen atom, and the macroscopic in terms of the ratio of the mass of the substance in gaseous form to the mass of O_2 which occupies the same volume at the same temperature and pressure. In each case the quantity is defined as a ratio.

In conformity with our use of the word relative the definitions suggest for this quantity the name *relative molar mass* (or relative molar weight).

9. Velocity.

It is frequently desired to express velocity in terms of the velocity of light, but the use of the term relative velocity for the ratio of the

* If molar mass is regarded as too pedantic then molar weight must be agreed to, but it is time we eliminated such loose terminology from physics.

velocity of a body to the velocity of light, though desirable for the sake of uniformity with the suggestions in this paper, is not permissible because of its entirely different use.

This raises the question as to whether *relative* is the most suitable word to adopt for the purpose in hand. The accepted meaning of *relative* with respect to velocities denotes the mathematical operation of subtraction, whereas throughout this paper it denotes the operation of division.

An alternative that suggests itself in connection with our derived quantities is the word *specific*. It does not recommend itself so readily, partly because of its association in rather different senses with such terms as specific resistance, specific heat, specific volume, specific charge; but this is not an overriding objection as there are either already in use or suggested in this paper alternatives to the first three, and the specific charge is not an essential term.

TABLE I.

<i>Present Terminology.</i>	<i>Suggested Terminology.</i>
Density.	Density.
Relative density.	Relative density.
Specific volume.	Rarity.
Specific heat.	Thermity.
	Relative thermity.
Dielectric constant.	Permittivity.
	Relative permittivity.
Permeability.	Permeability.
	Relative permeability.
Gram molecule (physical quantity).	Molar mass.
Gram molecule (unit).	Mole.
Molecular weight.	Relative molar mass.
Avogadro's number.	Avogadron.

In the first column of Table I. is given the current name for the various quantities discussed here, and in column two are shown the names suggested in this paper.

SYMBOLS.

10. Discussion.

The present position with regard to symbols is not very satisfactory. In this respect three problems arise:

(a) The use of different symbols in the various languages for the same physical quantity.

(b) The use of different symbols for the same quantity in English literature.

(c) The use of one symbol to denote several different quantities because of the limited number of symbols commonly employed.

We shall concern ourselves here with the more restricted, though by no means insignificant, problems involved in (b) and (c) and more particularly with (c).

The problem of the use of symbols has been attacked to a limited extent by the Joint Committee of the Chemical Society, the Faraday Society, and the Physical Society in its Report on Symbols for Thermodynamical and Physico-Chemical Quantities and Conventions relating to their use*. The purpose of the Committee was to choose for each particular physical quantity one symbol from those already in current use with a view to attaining consistency and avoiding to some extent the excessive use of a particular symbol for several quantities. This latter aim, however, cannot be achieved without a drastic change in certain cases, and moreover, some of the difficult cases were entirely omitted. However, to achieve any such change as would be necessary a much more authoritative and representative body is required comprising both American and British representatives.

The difficulties arising through the lack of sufficient number of symbols is becoming serious and is leading, at times, to confusion. The difficulty could be met to a considerable extent by introducing further symbols into general usage.

But before proceeding further it should be emphasized that there is no intention to suggest that one symbol and one symbol only should be used in all circumstances for a particular quantity. For instance, it would be absurd to suggest that P, Q, R should not be employed to represent forces in statics; or that P, Q, R, S should not denote resistances in bridge work. In such work there is no possibility of confusion; and many other similar examples could be cited where letters may be chosen almost at random to represent particular quantities. But in the presentation of more formal and general work both in textbooks and original papers there is a need to preserve as far as possible one symbol, at least, for each physical quantity.

This need not be carried to extremes. For example, p , representing momentum in generalized co-ordinates, is so obvious in its context that there would be little possibility of confusing it with p for pressure; again, l and L representing the latent heats are readily distinguishable from the quantum numbers l and L ; and so on for other quantities. Nevertheless, whenever the chance of confusion exists, and particularly when the quantities are liable to occur together, then it is most desirable that separate symbols should be adopted. A case in point concerns the quantity of heat and electric charge, both of which are often denoted by Q . Now these two quantities occur together in thermoelectricity, so that it is desirable to adopt two different symbols throughout rather than to change one of them just for thermoelectricity. It is to this type of problem that this paper is addressed.

11. *Illustrations of the Excessive Use of Some Symbols.*

Firstly, then, let us consider a few of the symbols which are already used to denote two or more quantities and which thereby may lead to

* Referred to in the following as the Symbols Committee,

confusion. To take some examples, the letters M , m , μ , N , n , I each denote several physical quantities.

To use M and m to represent mass only raises sufficient difficulties, even with the introduction of suffixes, without having to employ the same symbols for quantities such as magnetic moment, magnetic pole strength, and magnetic quantum number. Occasionally μ is invoked as a symbol for mass to relieve the pressure on M and m , but this is not always convenient as μ itself is used for a number of physical quantities.

For example, the mass of the electron and the magnetic quantum number frequently occur together, and the difficulty is fully appreciated when we find more than one text-book writing

$$\Delta E = \frac{Hehm}{4\pi mc},$$

where m in the numerator denotes the magnetic quantum number and m in the denominator the mass of the electron. μ may replace the mass m in such cases, but it is just in magnetic problems that μ is already used for two quantities—permeability, and magnetic moment especially of atoms and molecules.

N and n are used to denote numbers, numbers per unit length, numbers per unit volume, and occasionally other physical quantities. Often one requires a further symbol in this connexion without having recourse to suffixes. Sometimes ν serves this purpose, but it is so universally employed for both frequency and wave number that it is undesirable in many cases to use it for a pure number as well. Moreover, apart from the shortage of symbols in this case, there is no accepted practice for distinguishing between a pure number and a number per unit length or per unit volume.

I is another symbol which presents some difficulty. When it represents intensity of light, little difficulty arises because it is seldom as such associated with other quantities denoted by I . But moment of inertia, electric current, and intensity of magnetization, for all of which I is used, do occur together.

12. Conditions to be Satisfied when Selecting Symbols.

Sufficient then has been said to indicate the problem. A partial, but only a partial solution may be obtained by the copious use of suffixes. These are, at best, ungainly and in any case numerical suffixes should be adopted only to modify a quantity and not to change its entire meaning. For example, m_1 and m_2 should only denote two different masses and not m_1 a mass and m_2 a quantum number. On the other hand, a case can be made out for the occasional use of letter suffixes to denote a different physical quantity. c denotes the velocity of light, but when the suffix p or v is added—e.g. c_p and c_v —then the symbol denotes a specific heat. Even this is not satisfactory, for when we come to solids and liquids the p and v have lost their significance and are omitted. But as c without a suffix cannot be used for a specific heat,

then the notation has to be changed and s is generally used—another point, incidentally, with which the Symbols Committee did not deal. In general, then, suffixes should not be resorted to in order to distinguish between different physical quantities.

A further limiting factor in the choice of symbols is that occasionally it is desirable to retain both the capital and the lower-case letter for the same physical quantity, such as both M and m for mass.

The most desirable solution is the introduction, by some representative and authoritative body, of more symbols into physics. Let us consider then the conditions that any symbol in physics should satisfy.

- (1) When printed, it should be easily distinguished from other symbols.
- (2) It should lend itself readily to representation in manuscript and give no opportunity of confusion with other symbols.
- (3) It should have a distinguishing name.

Of the three conditions the second is the most difficult to satisfy. For the first condition many variations on the one letter can be made by the use of different printing types, but most of these are not easily translated into readily distinguishable manuscript forms.

This is a serious objection to the Symbols Committee's recommendation of the use of black italics for certain constants; for example J for mechanical equivalent of heat, m for the rest mass of the electron, e for the electronic charge. The various methods, such as underlining, for denoting particular types may be suitable in a manuscript to be translated into type by a compositor, but for ordinary manuscript and blackboard work the method does not recommend itself. The Symbols Committee recommended that these symbols printed in black italics could be represented in manuscript by a cursive or script letter. But this can only be applied to capitals. Of the ten symbols for which black italics were recommended only four are capitals, leaving six lower-case letters which cannot be readily translated into manuscript in the way suggested. Moreover, the use of heavy type so readily suggests vectors, and the fact that one may be in italics and the other in Clarendon type does not seem to be sufficient justification for the suggestion.

The third condition does not raise any real difficulty and will be referred to later.

The symbols now in general use are the lower-case letters of the English and Greek alphabets, the English capitals, and the Greek capitals which differ from the English. Other than these, the suggestions divide themselves into four groups: (a) English script capitals, (b) Inverted letters, (c) Letters from alphabets other than English and Greek, (d) Composite and synthetic symbols. None of these are new suggestions but their possibilities have not been explored to any considerable extent.

(a) *English Script Capitals*.—The method of using script as a variant is fairly common, but is limited to a few letters and is generally resorted to only when notation difficulties arise in a particular piece of work, so that there is no universal use of script letters to denote particular

physical quantities. Nearly all the English script capitals are quite distinctive both in print and manuscript, because in both cases the ordinary capitals are represented by italic or Roman type.

(b) *Inverted Letters* *.—This second suggestion is a much more drastic innovation. Many of the letters treated in this manner give quite a distinctive symbol. One point should be noticed in this respect; when a capital is inverted part of the letter falls below the line of type. Thus when E and R are inverted we have \mathfrak{E} and \mathfrak{R} . This may be considered a disadvantage, but such letters would not upset the aligning of type as they would extend no further below the line of type than p , y , etc. Moreover, it would further distinguish such a symbol from the same letter printed in its normal position.

The more general use of script capitals and the introduction of inverted letters should provide most of the extra symbols required, but if more are needed either as alternatives or supplements to those already put forward in (a) and (b), then (c) and (d) are possibilities.

(c) *Alphabets other than English and Greek*.—There are several alphabets which would lend themselves readily to this purpose, but objections may be raised firstly that such alphabets are unfamiliar and secondly that many printers have not the type available. After English and Greek the alphabets which suggest themselves are Russian and Hebrew. The Russian, and to a less extent the Hebrew, alphabets contain a number of letters which would be suitable as symbols. But this is by no means the extent of our choice. For example, the Georgian and Armenian alphabets are a most fertile field in this respect and have some most attractive and distinctive letters, but a serious objection is the certainty in these two cases that the type would have to be cut especially for the purpose.

(d) *Composite and Synthetic Symbols*.—This suggestion has the disadvantages of unfamiliarity of the symbols, the need for cutting new type for the purpose, and the coining of new names, although some such symbols have been introduced into physics, such as Dirac's \hbar . As examples of some composite symbols we might have b and p superposed, d and p superposed, or a composite letter of h and m . Another possibility is the use of double letters such as fi , ff , as such pairs of letters are cast in single pieces when setting up type.

13. Naming of Symbols.

The question of the names for the symbols need not involve us in much discussion. Already we have to apply the adjectives *capital* or *small* to a letter when it is necessary to distinguish between the upper and lower case, and therefore, it is only a step further in the same direction to apply the adjectives *script* and *inverse* respectively to a

* This device has already been used to a limited extent for a rather different purpose by Whitehead and Russell ⁽³⁾. Also the inverted delta, ∇ , is already in general use as a mathematical symbol.

script and an inverted letter. If Russian, Hebrew or any other type of letters were used, many of these have a name differing from that applied to the corresponding English or Greek letter.

14. *Modifying Signs.*

As already mentioned, the use of suffixes and other modifying signs should in general be employed only in conjunction with a given physical quantity and not to denote an entirely separate quantity. This applies particularly to the use of numerals as modifying signs.

Apart from the customary use of indices and suffixes there is a further possibility. Some writers wish to distinguish between electric and magnetic symbols according to whether they are expressed in electrostatic, electromagnetic or practical units. Sufficient symbols are not available to permit the adoption of, say, lower case, capitals, and script capitals for this purpose, but the introduction of indices or suffixes in this case is quite justified. There are several possibilities. Let us take electric potential as an example; expressed in practical units, e.s.u. and e.m.u., the potential may be represented respectively as V , V^* , V^\dagger . These two signs have the disadvantage that they often refer to footnotes, and in addition to this the asterisk may denote the complex conjugate of a quantity. An alternative set of symbols would be V , V^a , V^b . If suffixes are preferred to indices then we might adopt V , V_s , V_m .

15. *Conclusion.*

As stated earlier, the purpose of this paper is to suggest that physicists should, through a representative and authoritative committee, come to some agreement on the terminology and symbols to be used in physics, paying special attention to the need for clearing up some confusion in the terminology, and to the possibility of introducing further symbols into general use.

At a time when there is so much emphasis on the application of scientific methods to social and economic planning, is it too much to hope that scientists may also see the need for putting their own house in order and take steps accordingly to deal with this problem of terminology and symbols?

In conclusion, the author would like to acknowledge his indebtedness to a number of colleagues and friends with whom he has had the opportunity of discussing these problems and who have made a number of suggestions.

References.

- (1) Duncanson, *Proc. Phys. Soc.* liv. p. 504 (1942).
- (2) Guggenheim, *Phil. Mag.* xxxiii. p. 479 (1942).
- (3) Whitehead and Russell, 'Principia Mathematica,' C.U.P. (1910).

University College, London,
c/o University College of North Wales,
Bangor,

XI. The Geometry of Demagnetizing Curves, and its Consequences for Permanent Magnet Materials.

By K. HOSELITZ, Ph.D.*

[Received May 5, 1943.]

Introductory.

It has been pointed out by various authors (Evershed, 1920, Sanford, 1927, Hornfeck and Edgar, 1940) that for the design of permanent magnets a knowledge of the demagnetizing curve of the magnet material is necessary. By the term "demagnetizing curve" we mean the descending part of the hysteresis loop which lies between the remanence ($B=B_r$, $H=0$) and the coercive force ($B=0$, $H=H_c$) †.

Elementary considerations show that the total available energy in the gap of a permanent magnet $H_1^2 v = \eta \cdot BH \cdot V$ where H_1 is the field in the gap, v the gap volume, B the flux density in the magnet, H the field in the magnet, and V the magnet volume. $1/\eta$ is the leakage factor which will approach unity as the gap length decreases to zero. The most economical use of permanent magnet material is obtained where the product BH is a maximum. It is essential, therefore, to know the variation of BH with B or H , so that the maximum value of this product may be chosen. This point $(BH)_{\max}$, together with the two intercepts on the co-ordinate axes, B_r and H_c , constitute the three features of the demagnetizing curve that are most essential to permanent magnet designers.

A well-known empirical feature of the demagnetizing curve leads to a simplified treatment. The $(BH)_{\max}$ point lies very close to the intersection of the demagnetizing curve with the diagonal of a rectangle, which has one corner at the origin, and two other corners at the points B_r and H_c . This characteristic is followed fairly closely in alloys where H_c is numerically considerably smaller than B_r . For materials in which H_c is of the order of magnitude of B_r , the demagnetizing curve approaches more closely to a straight line, and in such cases the $(BH)_{\max}$ point coincides exactly with the postulated point. Thus, since this approximation is experimentally justified, we will assume with Watson (1923) and others, that it is a necessary characteristic of the demagnetizing curve, for which no full theoretical explanation is at present available. Some of our arguments are based on this assumption.

* Communicated by the Author.

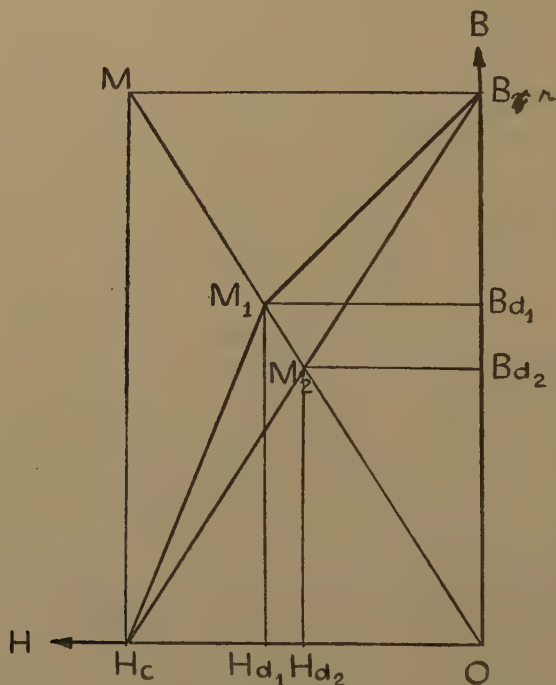
† In the following paper, the negative sign of H_c is omitted for convenience. Attention is drawn to the fact that the flux density B is plotted, and not the magnetization $4\pi I$, which is of no consequence for permanent magnets.

In the following, an attempt is made to give some further information about demagnetizing curves by describing methods of a purely geometrical character, which allow a number of general theoretical deductions about B-H curves to be made.

Limiting Demagnetizing Curves.

We can find limiting demagnetizing curves geometrically, by postulating a number of assumptions which are reasonable and seem to be generally applicable. The following treatment is applied to a permanent magnet material with given values of H_c and B_r . It is assumed that

Fig. 1.



the $(BH)_{\max}$ point lies exactly at the intersection of the demagnetizing curve with the diagonal OM of rectangle OH_cMB_r (fig. 1). This condition is closely fulfilled in all known materials, and the generalization to all possible materials is justified, as has been mentioned above. We also assume that the demagnetizing curve is not concave towards the top, a condition which is also fulfilled by all known materials. With these assumptions as basis, we find that for given values of B_r and H_c there are two limiting demagnetizing curves, one with a maximum value of $(BH)_{\max}$, the other with a minimum value.

The former may be found by postulating a material whose magnetization I is constant over a certain range beyond the point $H=0$. This corresponds to a B-H curve in this region with a slope of unity, since

$B = H + 4\pi I$, and $4\pi I$ is constant. The B - H curve will intersect the diagonal OM of the rectangle OH_cMB_r at the point M_1 . From this point B decreases at a greater rate, to zero at H_c , and it is therefore evident that M_1 corresponds to the maximum value of BH , *i. e.* $(BH)_{\max}$. We are concerned only with the area of the rectangle $OH_dM_1B_{d_1}$, and it is irrelevant how the B - H curve continues between M_1 and H_c . Our fictitious B - H curve so constructed fulfils all our assumptions, in that it passes through the given points B_r and H_c ; its $(BH)_{\max}$ point M_1 lies on the diagonal OM of the rectangle OH_cMB_r ; and it is nowhere concave towards the top. Further, it can be seen that no B - H curve with a higher value of $(BH)_{\max}$ can possibly exist which obeys our assumptions, because any demagnetizing curve which cuts the diagonal OM at a point nearer to M than M_1 , and passes through B_r , must necessarily have a slope of less than unity somewhere between these two points. But if this were possible it would mean that the magnetization increases with falling magnetizing field, as can be seen at once :

$$B = H + 4\pi I.$$

$$\therefore \frac{dB}{dH} = 1 + 4\pi \frac{dI}{dH}, \text{ and if } \frac{dB}{dH} < 1, \frac{dI}{dH} < 0,$$

i. e., I rises with falling H , a condition which is obviously nonsensical.

We will now define a factor γ , called the fullness factor or curve factor, which is simply a measure of the squareness or fullness of any given demagnetizing curve :

$$\gamma = \frac{B_d \cdot H_d}{B_r \cdot H_c};$$

i. e., it is the area of the rectangle with the co-ordinates of the $(BH)_{\max}$ point as sides, divided by the area of the rectangle with B_r and H_c as sides. For any given values of B_r and H_c , the material with the higher $(BH)_{\max}$ value will have a greater value of γ , that is, a "fuller" demagnetizing curve. Since the curve constructed above has the highest value of $(BH)_{\max}$ for a given B_r and H_c , hence the highest γ , we will call it the γ_{\max} curve, and the corresponding value of $(BH)_{\max}$ we call $(BH)_{\max}\gamma_{\max}$. It can be seen that $(BH)_{\max}\gamma_{\max}$ is a simple function of H_c and B_r . The co-ordinates of M_1 , the $(BH)_{\max}\gamma_{\max}$ point, are B_{d_1} and H_{d_1} . Since the slope of $\overline{B_rM_1}$ is unity,

$$\overline{B_rB_{d_1}} = \overline{M_1B_{d_1}}; \quad \overline{B_rB_{d_1}} = H_{d_1}; \quad B_{d_1} = B_r - H_{d_1};$$

and since M_1 is on the diagonal OM ,

$$\frac{OH_{d_1}}{OB_{d_1}} = \frac{H_c}{B_r}; \quad \text{or} \quad \frac{H_{d_1}}{B_{d_1}} = \frac{H_c}{B_r}.$$

This gives

$$B_{d_1} \cdot H_{d_1} = (BH)_{\max}\gamma_{\max} = \frac{B_r^3 H_c}{(B_r + H_c)^2}.$$

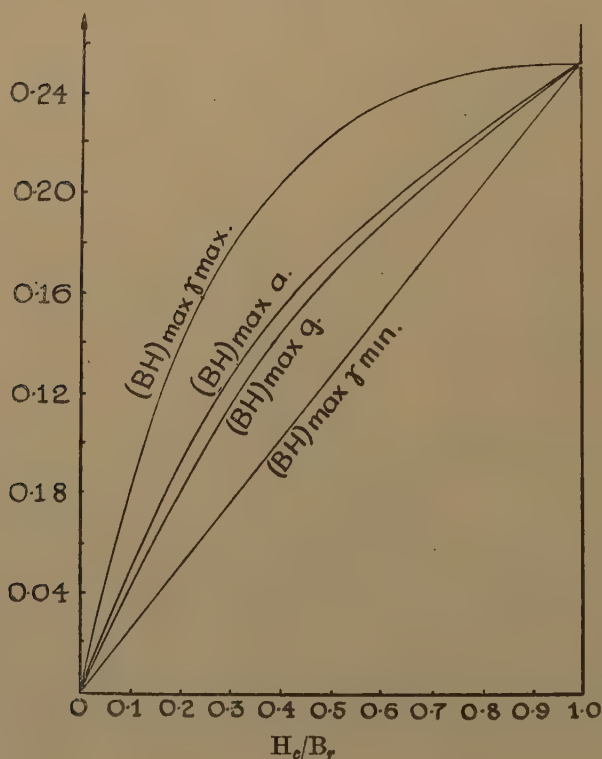
It will be of advantage if we express all values as functions of B_r and

$\frac{H_c}{B_r}$, because this allows the comparison of materials with different sets of values of B_r and H_c . In this notation we have, therefore,

$$(BH)_{\max} \gamma_{\max} = B_r^2 \times \frac{\frac{H_c}{B_r}}{\left(1 + \frac{H_c}{B_r}\right)^2} \quad \dots \quad (1)$$

It is known that since the slope of the B-H curve cannot be less than 1, H_c can never be greater than B_r , and hence $\frac{H_c}{B_r}$ will have values

Fig. 2.



between 0 and 1 only, and we can plot $(BH)_{\max} \gamma_{\max}$ in units of B_r^2 against $\frac{H_c}{B_r}$ (fig. 2). An interesting consequence of formula (1) is the fact that

$$\gamma_{\max} = \frac{(BH)_{\max} \gamma_{\max}}{B_r \times H_c} = \frac{B_r^2}{(B_r + H_c)^2},$$

and this function of $\frac{H_c}{B_r}$ becomes

favourable case the fullness decreases with increasing $\frac{H_c}{B_r}$, and no demagnetizing curve can have a fullness γ of less than $\frac{1}{4}$. We shall see below that these conclusions have important practical implications.

Practical Demagnetizing Curves.

In practice the demagnetizing curves of a material will not coincide with either of the two extreme possibilities of fig. 1, but will lie somewhere between the two. Therefore, two means between the two limiting cases will be considered, their choice being justified below. The two intermediate possibilities which are selected as approximations to practical B-H curves so far as their $(BH)_{\max}$ values as functions of B_r and H_c are concerned, are the geometric and arithmetic means of the two extreme curves.

In the previous paragraph we have expressed $(BH)_{\max}$ of the two limiting cases as functions of B_r and H_c , and by forming the geometric mean, $(BH)_{\max g}$, and the arithmetic mean, $(BH)_{\max a}$, of these two extremes, we can express the practical values as functions of B_r and H_c , or more conveniently of B_r and $\frac{H_c}{B_r}$.

Thus the geometric mean,

$$(BH)_{\max g} = B_r^2 \times \frac{\frac{H_c}{B_r}}{2 \left(1 + \frac{H_c}{B_r} \right)}, \quad \dots \dots \dots (5)$$

and the corresponding fullness factor :

$$\gamma_g = \frac{1}{2 \left(1 + \frac{H_c}{B_r} \right)}, \quad \dots \dots \dots (6)$$

and similarly the arithmetic mean,

$$(BH)_{\max a} = B_r^2 \times \frac{\frac{H_c}{B_r} \left[\left(\frac{H_c}{B_r} \right)^2 + 2 \frac{H_c}{B_r} + 5 \right]}{8 \left(1 + \frac{H_c}{B_r} \right)^2}, \quad \dots \dots (7)$$

and the fullness factor :

$$\gamma_a = \frac{\left(\frac{H_c}{B_r} \right)^2 + 2 \frac{H_c}{B_r} + 5}{8 \left(1 + \frac{H_c}{B_r} \right)^2} \dots \dots \dots (8)$$

The reason why the geometric and arithmetic means are chosen for comparison with practical cases is that the $(BH)_{\max}$ values of ordinary magnet alloys are very close to these two simple functions. The $(BH)_{\max}$ of ordinary isotropic alloys of the FeNiAl type corresponds

very closely to $(BH)_{\max g}$, while anisotropic materials of this type correspond more closely to $(BH)_{\max a}$. Two examples are given to illustrate this agreement.

1. *Isotropic Material*.—For Alnico the observed values are

$$B_r = 8000, \quad (BH)_{\max} = 1.7 \times 10^6, \quad H_c = 500, \quad \frac{H_c}{B_r} = 0.625.$$

The calculations give

$$(BH)_{\max g} = 1.88 \times 10^6. \quad (\text{Formula 5.})$$

$$(BH)_{\max a} = 2.27 \times 10^6. \quad (\text{Formula 7.})$$

2. *Anisotropic Material*.—For Alcomax the observed values are

$$B_r = 12500, \quad (BH)_{\max} = 3.5 \times 10^6, \quad H_c = 500, \quad \frac{H_c}{B_r} = 0.04.$$

By calculation, $(BH)_{\max a} = 3.66 \times 10^6$. (Formula 7.)

$$(BH)_{\max g} = 3.0 \times 10^6. \quad (\text{Formula 5.})$$

In other cases of permanent magnet alloys in use, similar agreement is obtained, and it seems that the choice of these two means as a guide is not unjustifiable. One would expect that the $(BH)_{\max}$ values of isotropic and anisotropic materials should approximate to different functions of B_r and H_c , since the two cases can be represented by different arrangements and distribution of the elementary magnets concerned. A closer study of the theoretical conditions underlying these relationships is under consideration.

The values of $(BH)_{\max g}$ and $(BH)_{\max a}$ are plotted as functions of $\frac{H_c}{B_r}$ in units of B_r^2 in fig. 2, and the corresponding values of γ_g and γ_a in figs. 4 and 5 respectively. In both cases it can be seen that the fullness factor of a demagnetizing curve decreases with increasing $\frac{H_c}{B_r}$. This would have been expected from any curve intermediate between the two limits, because γ_{\max} has this characteristic, and γ_{\min} is constant, and hence any curve between these two must necessarily fall from higher values for small $\frac{H_c}{B_r}$ to the only possible value at $\frac{H_c}{B_r} = 1$, namely $\gamma = \frac{1}{2}$.

As a consequence, any magnet alloy with a high ratio of $\frac{H_c}{B_r}$, or in other words with a high coercive force relative to the remanence, must be somewhat inefficient on account of its demagnetizing curve approaching a straight line connecting B_r and H_c . Such alloys are mentioned in the literature quite frequently, *e.g.* certain Heusler alloys (Persson, 1938, Messkin & Kussmann, 1932), or recently some alloys of platinum with a ferromagnetic element, such as iron or cobalt (Graf & Kussmann, 1935, Jellinghaus, 1937). Some of these have a ratio of $\frac{H_c}{B_r}$ approaching unity,

and in spite of the fact that B_r and H_c of several thousand c.g.s. units are recorded, their $(BH)_{\max}$ values are not outstanding compared with those of modern high performance alloys of the FeNiAl type. This disappointing performance of alloys with remarkably high figures, especially of H_c , is due to the necessarily low values of γ (the demagnetizing curves are nearly straight lines), and this fact cannot be altered by any heat-treatment as it is a consequence of a physical condition, *i e.*, that the magnetization cannot rise with decreasing field.

Discussion.

As mentioned before, it can be seen from the graphs in fig. 2 that the $(BH)_{\max}$ values of magnet materials increase with increasing values of $\frac{H_c}{B_r}$, and this increase is greater for low values of $\frac{H_c}{B_r}$, the slope decreasing for $\frac{H_c}{B_r}$ approaching unity. Since the $(BH)_{\max}$ values are plotted in units of B_r^2 , it is obvious that they increase with B_r for any given H_c . These characteristics suggest that there may be a point of $\frac{H_c}{B_r}$ where a small increase of B_r yields a greater gain in $(BH)_{\max}$ than a small increase in H_c . For low values of $\frac{H_c}{B_r}$ where the slope of $(BH)_{\max}$ is great, it will be of greater advantage to increase H_c in order to raise $(BH)_{\max}$. Beyond a certain value of $\frac{H_c}{B_r}$ a further increase in H_c will, however, be less favourable than a small increase in B_r . This can be shown by calculating the values of

$$\left. \frac{\partial(BH)_{\max}}{\partial H_c} \right|_{B_r = \text{const.}} \quad \text{and} \quad \left. \frac{\partial(BH)_{\max}}{\partial B_r} \right|_{H_c = \text{const.}}$$

and plotting them as functions of $\frac{H_c}{B_r}$. If, for a given function $(BH)_{\max}$ of B_r and H_c the two partial differential coefficients plotted against $\frac{H_c}{B_r}$ intersect, the corresponding value of $\frac{H_c}{B_r}$ is the one for which an increase in B_r increases $(BH)_{\max}$ just as much as an equal increase in H_c . At lower values of $\frac{H_c}{B_r}$ it is better to increase H_c , at higher values of $\frac{H_c}{B_r}$ it is better to increase B_r .

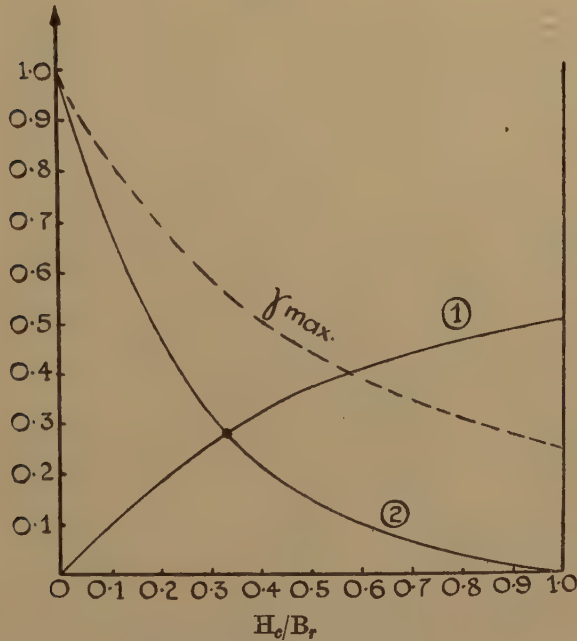
These differential coefficients have been calculated, and are as follows :—

$$\left. \frac{\partial(BH)_{\max}}{\partial B_r} \right|_{H_c = \text{const.}} = B_r \times \frac{\frac{H_c}{B_r} \left(1 + 3 \frac{H_c}{B_r} \right)}{\left(1 + \frac{H_c}{B_r} \right)^3} \quad (9a)$$

$$\left. \frac{\partial(BH)_{\max} \gamma_{\max}}{\partial H_c} \right|_{B_r = \text{const.}} = B_r \times \frac{\left(1 - \frac{H_c}{B_r}\right)}{\left(1 + \frac{H_c}{B_r}\right)^3} \quad (9b)$$

(plotted in fig. 3 in units of B_r together with γ_{\max} (dotted)).

Fig. 3.



$$\left. \frac{\partial(BH)_{\max} \gamma_{\max}}{\partial B_r} \right|_{H_c = \text{const.}}; \text{ (formula 9 a).} \quad (1)$$

$$\left. \frac{\partial(BH)_{\max} \gamma_{\max}}{\partial H_c} \right|_{B_r = \text{const.}}; \text{ (formula 9 b).} \quad (2)$$

$$\left. \frac{\partial(BH)_{\max} g}{\partial B_r} \right|_{H_c = \text{const.}} = B_r \times \frac{\left(\frac{H_c}{B_r}\right) \left(1 + 2 \frac{H_c}{B_r}\right)}{2 \left(1 + \frac{H_c}{B_r}\right)^2} \quad (10a)$$

$$\left. \frac{\partial(BH)_{\max} g}{\partial H_c} \right|_{B_r = \text{const.}} = B_r \times \frac{1}{2 \left(1 + \frac{H_c}{B_r}\right)^2} \quad (10b)$$

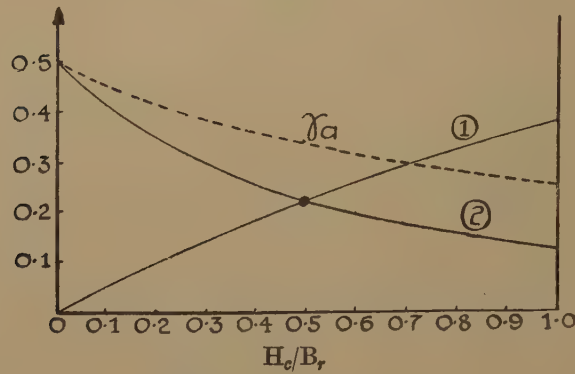
(plotted in fig. 4 in units of B_r together with γ_g (dotted)).

$$\left. \frac{\partial(BH)_{\max} a}{\partial B_r} \right|_{H_c = \text{const.}} = B_r \times \frac{5 \left(\frac{H_c}{B_r}\right) + 15 \left(\frac{H_c}{B_r}\right)^2 + 3 \left(\frac{H_c}{B_r}\right)^3 + \left(\frac{H_c}{B_r}\right)^4}{8 \left(1 + \frac{H_c}{B_r}\right)^3} \quad (11a)$$

$$\left. \frac{\partial(BH)_{\max a}}{\partial H_c} \right|_{B_r = \text{const.}} = B_r \times \frac{5 - \left(\frac{H_c}{B_r}\right) + 3 \left(\frac{H_c}{B_r}\right)^2 + \left(\frac{H_c}{B_r}\right)^3}{8 \left(1 + \frac{H_c}{B_r}\right)^3} \quad (11b)$$

(plotted in fig. 5 in units of B_r together with γ_a (dotted)).

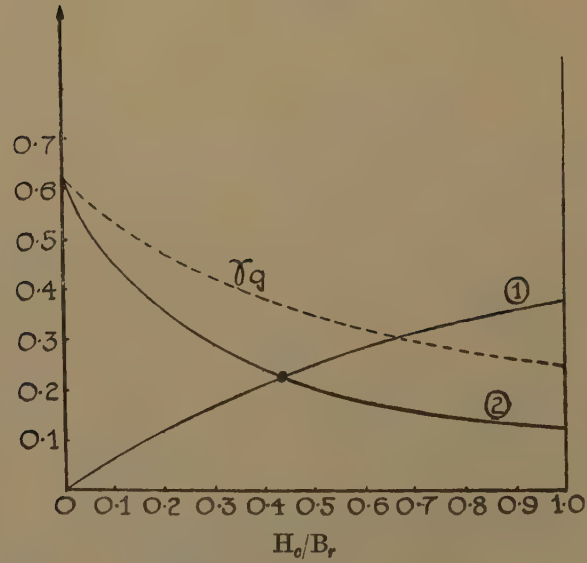
Fig. 4.



$$\frac{\partial(BH)_{\max g}}{\partial B_r}; \text{ (formula 10 a).} \quad (1)$$

$$\frac{\partial(BH)_{\max g}}{\partial H_c}; \text{ (formula 10 b).} \quad (2)$$

Fig. 5.



$$\frac{\partial(BH)_{\max a}}{\partial B_r}; \text{ (formula 11 a).} \quad (1)$$

$$\frac{\partial(BH)_{\max a}}{\partial H_c}; \text{ (formula 11 b).} \quad (2)$$

Inspection of the curves for $\frac{\partial(BH)_{\max}}{\partial B_r}$ and $\frac{\partial(BH)_{\max}}{\partial H_c}$ for the three functions shows that in each case the two curves intersect at a certain value of $\frac{H_c}{B_r}$, which, as we have seen above, is the critical ratio of $\frac{H_c}{B_r}$, where further increase in H_c is actually of less advantage than an increase in B_r . This value can be seen either from the graphs, or by equating the corresponding functions :

For the "fullest" curve $\frac{\partial}{\partial B_r} = \frac{\partial}{\partial H_c} \therefore \frac{H_c}{B_r} = \frac{1}{3},$

For the geometric mean $\frac{H_c}{B_r} = \frac{1}{2},$

For the arithmetic mean $\frac{H_c}{B_r} = 0.42.$

This means that for any alloy with $\frac{H_c}{B_r} < \frac{1}{3}$ it will in any case be more advantageous to increase H_c by a small amount than to increase B_r , similarly. This will still apply for most alloys, isotropic as well as anisotropic, so long as $\frac{H_c}{B_r}$ is below about 0.42-0.5. In the two cases quoted above, the gain from increasing H_c is considerably greater than the gain from increasing B_r , because $\frac{H_c}{B_r}$ is so far below the value of 0.42-0.5. Hence for commercial alloys of the FeNiAl type an increase in H_c with constant B_r is about 15-20 times as effective as *vice versa*. According to these conclusions the practical problem devolves into finding alloys with larger coercive forces rather than higher remanences, because our present magnetic materials have a $\frac{H_c}{B_r}$ ratio of not greater than 0.15. At the same time it should be borne in mind that increasing the coercive force alone is not economical beyond certain limits, which we tentatively set at about $H_c = 0.45 \times B_r$, without at least a corresponding increase in B_r .

Commercial alloys of the most modern anisotropic types have $(BH)_{\max}$ values approaching 5×10^6 in units of $\frac{1}{8}\pi$ ergs/cm³. This shows that no improvement in $(BH)_{\max}$ values is to be expected from any alloy with a remanence of 5000 gauss or less, however great its coercive force may be, because the greatest $(BH)_{\max}$ attainable from any alloy with $B_r = 5000$ is, according to the above considerations, $\frac{B_r^2}{4} = 6.25 \times 10^6$, and this is in the case of $H_c = B_r$, which is not in practice realizable. It seems, therefore, that the best way to develop alloys of higher $(BH)_{\max}$ than those at

present available is to try to obtain alloys with remanences such as the best alloys have at present, but with increased coercive forces. *

The writer wishes to express his thanks to Professor W. Sucksmith, F.R.S., for his continued interest and valuable suggestions, and to Dr. A. Edwards and Mr. J. E. Gould, of the Permanent Magnet Association, for helpful discussions during the course of this work.

References.

- Evershed, *Journal I.E.E.* vol. lviii. p. 780 (1920).
 Graf and Kussmann, *Phys. Z.* vol. xxxvi. p. 544 (1935).
 Hornfeck and Edgar, *A.I.E.E. Tech. Paper*, p. 40 (1940).
 Jellinghaus, *A.T.M.* vol. lxix (1937).
 Messkin and Kussmann, 'Die ferromagnetischen Legierungen' (1932).
 Persson, *Naturwissenschaften*, vol. xvi. p. 613 (1938).
 Sanford, *Bur. Standards Sci. Papers*, No. 567 (1927).
 Watson, *Journal I.E.E.* vol. lxi. p. 641 (1923).

Research Laboratory,
 Permanent Magnet Association,
 University of Sheffield.

XII. On a Problem in Heat Conduction.

By C. J. TRANTER, M.A.*

[Received September 8, 1943.]

IN a recent paper in this Journal ⁽¹⁾, Mr. A. N. Lowan finds the steady temperature in the semi-infinite annulus bounded by the plane $z=0$ and the cylindrical surfaces $r=a$, $r=b$ when the surfaces $z=0$, $r=a$ are maintained at zero temperature and the surface $r=b$ at temperature $f(z)$. The problem is solved by the classical method of separating the variables in the differential equation and using the well-known double integral representation for $f(z)$.

A formal solution of this problem can be obtained very easily by the use of a Fourier transform, and the more difficult problem of the variable temperature can be solved formally by using a Fourier transform followed by a Laplace transform. A solution on these lines is given below for the variable temperature in the case where the initial temperature of the annulus is zero.

For the variable temperature problem, the temperature T must satisfy

$$\frac{\partial^2 T}{\partial r^2} + \frac{1}{r} \frac{\partial T}{\partial r} + \frac{\partial^2 T}{\partial z^2} = \frac{1}{k} \frac{\partial T}{\partial t} \quad b < r < a, \quad z > 0, \quad t > 0, \quad \dots \quad (1)$$

where k is the diffusivity.

* Communicated by the Author.

The boundary and initial conditions are

$$T=0, \quad b < r < a, \quad t > 0, \quad z=0. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$T=0, \quad z > 0, \quad t > 0, \quad r=a. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$T=f(z), \quad z > 0, \quad t > 0, \quad r=b. \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$T=0, \quad b < r < a, \quad z > 0, \quad t=0. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Write
$$T_s = \sqrt{\frac{2}{\pi}} \int_0^\infty T \sin(\lambda z) dz. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Integrating by parts and using (2), we have

$$\sqrt{\frac{2}{\pi}} \int_0^\infty \frac{\partial^2 T}{\partial z^2} \sin(\lambda z) dz = -\lambda^2 T_s.$$

Hence, multiplying (1), (3), (4) and (5) by $\sqrt{\frac{2}{\pi}} \sin(\lambda z)$ and integrating with respect to z from 0 to ∞

$$\frac{\partial^2 T_s}{\partial r^2} + \frac{1}{r} \frac{\partial T_s}{\partial r} - \lambda^2 T_s = \frac{1}{k} \frac{\partial T_s}{\partial t}, \quad b < r < a, \quad t > 0. \quad . \quad . \quad . \quad (7)$$

$$T_s = 0, \quad t > 0, \quad r = a. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$T_s = \sqrt{\frac{2}{\pi}} \int_0^\infty f(z') \sin(\lambda z') dz', \quad t > 0, \quad r = b. \quad . \quad . \quad . \quad (9)$$

$$T_s = 0, \quad b < r < a, \quad t = 0. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

The transformation (6) has reduced the number of independent variables from three to two. A further reduction can be made by writing

$$\bar{T}_s = \int_0^\infty e^{-pt} T_s dt, \quad p > 0. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Integrating by parts and using (10)

$$\int_0^\infty e^{-pt} \frac{\partial T_s}{\partial t} dt = p \bar{T}_s.$$

Hence, multiplying (7), (8) and (9) by e^{-pt} and integrating with respect to t from 0 to ∞

$$\frac{d^2 \bar{T}_s}{dr^2} + \frac{1}{r} \frac{d \bar{T}_s}{dr} - \left(\lambda^2 + \frac{p}{k} \right) \bar{T}_s = 0, \quad b < r < a. \quad . \quad . \quad . \quad (12)$$

$$\bar{T}_s = 0, \quad r = a. \quad . \quad . \quad . \quad . \quad . \quad (13)$$

$$\bar{T}_s = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{p} \int_0^\infty f(z') \sin(\lambda z') dz', \quad r = b. \quad . \quad . \quad . \quad (14)$$

The solution of (12), satisfying (13) and (14), is

$$\bar{T}_s = \frac{F(a, b, r, \sqrt{\lambda^2 + p/k})}{p} \sqrt{\frac{2}{\pi}} \int_0^\infty f(z') \sin(\lambda z') dz', \quad . \quad . \quad . \quad (15)$$

where

$$F(a, b, r, \lambda) = \frac{I_0(\lambda r) K_0(\lambda a) - I_0(\lambda a) K_0(\lambda r)}{I_0(\lambda b) K_0(\lambda a) - I_0(\lambda a) K_0(\lambda b)}. \quad . \quad . \quad . \quad (16)$$

Using the Mellin Inversion Theorem ⁽²⁾ to obtain T_s from \bar{T}_s , we have

$$T_s = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{\zeta t} F(a, b, r, \sqrt{\lambda^2 + \zeta/k}) \frac{d\zeta}{\zeta} \cdot \sqrt{\frac{2}{\pi}} \int_0^\infty f(z') \sin(\lambda z') dz'. \quad (17)$$

The line integral can be evaluated by transforming it into a closed contour and applying the calculus of residues ⁽³⁾. The poles are at $\zeta=0$ and $\zeta=-k(\lambda^2 + \alpha_s^2)$ where $\pm\alpha_s$, $s=1, 2, \dots$ are the roots of

$$J_0(\alpha b)Y_0(\alpha a) - J_0(\alpha a)Y_0(\alpha b) = 0. \quad (18)$$

The pole at $\zeta=0$ gives $F(a, b, r, \lambda)$.

To find the residue at $\zeta=-k(\lambda^2 + \alpha_s^2)$, we require the result

$$\begin{aligned} & \left[\zeta \frac{d}{d\zeta} \{I_0(\mu b)K_0(\mu a) - I_0(\mu a)K_0(\mu b)\} \right]_{\zeta=-k(\lambda^2 + \alpha_s^2)} \\ &= \left[\frac{\zeta b}{2k\mu} \{I_0'(\mu b)K_0(\mu a) - I_0(\mu a)K_0'(\mu b)\} \right. \\ & \quad \left. + \frac{\zeta a}{2k\mu} \{I_0(\mu b)K_0'(\mu a) - I_0'(\mu a)K_0(\mu b)\} \right]_{\zeta=-k(\lambda^2 + \alpha_s^2)}, \quad (19) \end{aligned}$$

where we have written for convenience,

$$\mu = \sqrt{\lambda^2 + \zeta/k}. \quad (20)$$

Now, when $\zeta=-k(\lambda^2 + \alpha_s^2)$,

$$\frac{I_0(\mu b)}{I_0(\mu a)} = \frac{K_0(\mu b)}{K_0(\mu a)} = \frac{J_0(\alpha_s b)}{J_0(\alpha_s a)} = \xi, \text{ (say).}$$

Using this result, (19) becomes

$$\begin{aligned} & \left[\frac{\zeta b \xi^{-1}}{2k\mu} \{I_0'(\mu b)K_0(\mu b) - I_0(\mu b)K_0'(\mu b)\} \right. \\ & \quad \left. + \frac{\zeta a \xi}{2k\mu} \{I_0(\mu a)K_0'(\mu a) - I_0'(\mu a)K_0(\mu a)\} \right]_{\zeta=-k(\lambda^2 + \alpha_s^2)} \\ &= \left[\frac{\zeta}{2k\mu^2} (\xi^{-1} - \xi) \right]_{\zeta=-k(\lambda^2 + \alpha_s^2)} = \frac{\lambda^2 + \alpha_s^2}{2\alpha_s^2} \cdot \frac{J_0^2(\alpha_s a) - J_0^2(\alpha_s b)}{J_0(\alpha_s a)J_0(\alpha_s b)}. \quad (21) \end{aligned}$$

Also,

$$\begin{aligned} & [I_0(\mu r)K_0(\mu a) - I_0(\mu a)K_0(\mu r)]_{\zeta=-k(\lambda^2 + \alpha_s^2)} \\ &= -\frac{\pi}{2} [J_0(\alpha_s r)Y_0(\alpha_s a) - J_0(\alpha_s a)Y_0(\alpha_s r)]. \quad (22) \end{aligned}$$

Thus,

$$\begin{aligned} T_s &= \left[F(a, b, r, \lambda) - \pi \sum_{s=1}^{\infty} e^{-k(\lambda^2 + \alpha_s^2)t} \cdot \frac{\alpha_s^2}{\lambda^2 + \alpha_s^2} \cdot \frac{J_0(\alpha_s a)J_0(\alpha_s b)}{J_0^2(\alpha_s a) - J_0^2(\alpha_s b)} \right. \\ & \quad \left. \times \{J_0(\alpha_s r)Y_0(\alpha_s a) - J_0(\alpha_s a)Y_0(\alpha_s r)\} \right] \cdot \sqrt{\frac{2}{\pi}} \int_0^\infty f(z') \sin(\lambda z') dz'. \quad (23) \end{aligned}$$

The reciprocal relation to (6) is ⁽⁴⁾

$$T = \sqrt{\frac{2}{\pi}} \int_0^\infty T_s \sin(\lambda z) d\lambda \quad (24)$$

and so, finally,

$$T = \frac{2}{\pi} \int_0^\infty \int_0^\infty F(a, b, r, \lambda) f(z') \sin(\lambda z) \sin(\lambda z') d\lambda dz' \\ - 2 \sum_{s=1}^\infty e^{-k\alpha_s z} \cdot \frac{\alpha_s^2 J_0(\alpha_s a) J_0(\alpha_s b)}{J_0^2(\alpha_s a) - J_0^2(\alpha_s b)} \{J_0(\alpha_s r) Y_0(\alpha_s a) - J_0(\alpha_s a) Y_0(\alpha_s r)\} \\ \times \int_0^\infty \int_0^\infty e^{-k\lambda z'} \cdot \frac{f(z') \sin(\lambda z) \sin(\lambda z')}{\lambda^2 + \alpha_s^2} d\lambda dz'. \quad (25)$$

From the general solution (25), it is clear that the steady temperature can be obtained by allowing $t \rightarrow \infty$. If this alone is required, one transform only (the Fourier transform (6)) is sufficient.

References.

- (1) Lowan, Phil. Mag. (7) xxxiv. pp. 502-504 (1943).
- (2) Carslaw and Jaeger, 'Operational Methods in Applied Mathematics,' Oxford University Press, p. 71 (1941).
- (3) Carslaw and Jaeger, *loc. cit.* p. 75 *et seq.*
- (4) Titchmarsh, 'Introduction to the Theory of Fourier Integrals.' Oxford University Press, p. 4 (1937).

XIII. Introductory Notes on a Reformulation of the Special Theory of Relativity.

By P. A. SCOTT-IVERSEN, F.R.A.S.*

[Received July 7, 1943.]

Foreword by Professor H. DINGLE, Imperial College of Science and Technology.

This paper, containing a reformulation of the special theory of relativity, was submitted to me for criticism some months ago by Mr. P. A. Scott-Iversen, a very gifted physicist attached to the research staff of the Rover Company, Ltd. Its aim is to show that the relativity usually expressed by a modification of the measurements of length, time and mass when transformation is made to a moving system, may be equally well expressed by a variation of velocity with direction, while length, time and mass remain invariant. If this could be shown generally it would provide a re-expression of special relativity simpler for certain purposes than the orthodox theory. The author has, however, only shown that it is possible for two co-ordinate systems, though these may be any two between which the Lorentz transformation can normally be

* Communicated by Professor H. Dingle.

applied. It was his belief that this was sufficient for application to certain current problems of atomic physics, and I therefore suggested to him that it would be advisable to apply it to those problems, and so prove its utility, before seeking publication. Unfortunately, he died before being able to do this. The paper, therefore, with a few verbal alterations, is submitted for publication as a tribute to his memory and possibly as an indication of a new method of approach to the problems he had in view. At the least, it affords an interesting example of the variety of aspects in which the special theory of relativity may be regarded, and so makes a contribution to the better understanding of the essential principles of that theory.—H. D.

ABSTRACT.

A reformulation of the mechanics of the classical theory of special relativity in which, between any two inertial frames of reference, the units of length, time and mass are invariants, and simultaneity is absolute. The requirements of the Lorentz transformation are met by an appropriate choice of simultaneity for events at different places.

ONE of the postulates from which the transformation formulæ of the special theory of relativity are derived assures us that the velocity of light *in vacuo* is the same in all directions. Our belief in this statement is built on our knowledge of Roemer's astronomical measurements of the velocity of light and on the measurements carried out later by Fizeau and others. Roemer's measurement is usually regarded as giving the one-way velocity, while Fizeau's terrestrial type of measurement gives the there-and-back velocity.

It can, however, be shown that Roemer's method, although designed to give the one-way velocity, in fact, gives the there-and-back velocity in a manner quite similar to all terrestrial measurements of the velocity of light, and from this finding there can be evolved a new transformation system for the special theory of relativity in which the mechanical relations between any two inertial frames of reference in relative motion are expressed in invariant units of length, time and mass. This result influences our concept of simultaneity, which—unlike its form in the Einstein theory—now becomes absolute, that is: events that are simultaneous in one frame of reference are now simultaneous also in the other of the two frames of reference.

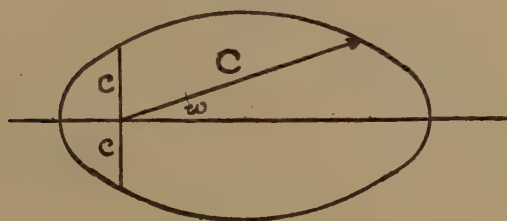
We shall introduce the new point of view by postulating a new distribution of the velocity of light, which is not the same in all directions. Then, by examples, we shall show that this new distribution in no way alters the result of Roemer's measurement or any other type of physical experiment involving velocities, and finally, we shall show that the particular distribution we have chosen forms a very simple basis for a new transformation system.

Let us assume that the there-and-back velocity measurements are accurate, and let us extend the scope of these measurements to cover any-way-there-and-any-way-back measurements: in other words, we postulate that the velocity of light *in vacuo* is always the same, when measured over any closed path. We also postulate that the one-way velocity of light in any medium is unknowable, and may be represented by an ellipsoidal distribution of velocities of light *in vacuo*, given in polar co-ordinates by the equation:

$$C = \frac{c}{1 - e \cos w} \quad \dots \dots \dots (1)$$

This equation is shown graphically in fig. 1, where the pole is one of the focal points, C is the velocity vector of light *in vacuo* (the one-way velocity), c is 3×10^{10} cm./sec. (the there-and-back velocity of light *in vacuo*), e is the eccentricity of the ellipsoid, the significance of which is to appear later, and w is the angle between the vector, C , and the major

Fig. 1.



axis of the ellipsoid. If $e=0$, the ellipsoid becomes a sphere, and the distribution becomes the one on which the classical special theory of relativity is based.

This covers only velocities of light *in vacuo*, and before we attack our problems we must find the corresponding distribution of smaller velocities, which also are unknowable as one-way velocities, since the distribution of simultaneity is determined by the unknowable one-way velocity of light. The distribution is found in the following manner.

A light signal, and a particle moving with a velocity lower than that of the signal, start simultaneously from one point on a given path and arrive, one after the other, at another point on the path. The time lapse between their arrivals is measured. This measurement must of course be the same irrespective of how we measure the velocities involved. Let us first regard the experiment under the assumption that the velocity distribution is spherical, that is, in accordance with the usual view. If the velocity of the particle is v , and the length of the path is l , then the time lapse is

$$t = \frac{l}{v} - \frac{l}{c} \quad \dots \dots \dots (2)$$

If, now, V is the vector we want to find, corresponding to C in (1), we substitute V and C for v and c in (2) and have

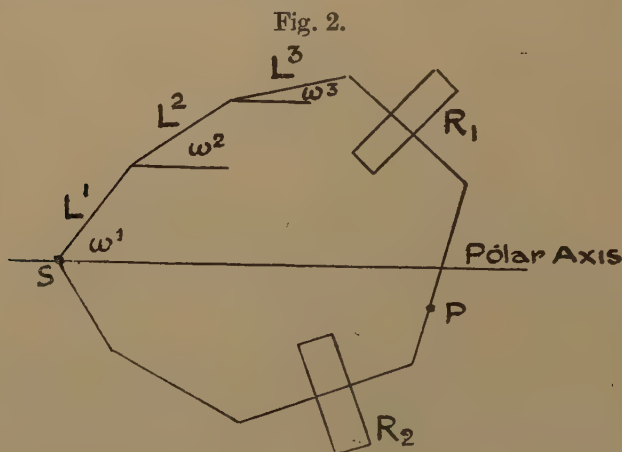
$$t = \frac{1}{v} - \frac{1}{c} = \frac{1}{V} - \frac{1}{C} = \frac{1}{V} - \frac{1}{c}(1 - e \cos w), \quad \dots \dots (3)$$

which gives

$$V = \frac{v}{1 - \frac{v}{c} e \cos w} \quad \dots \dots \dots (4)$$

This expression covers all velocities from rest to the velocity of light *in vacuo*, as may be seen by putting $v=c$, which gives (1).

We shall now show that this distribution of velocities, when applied to certain representative experiments, gives precisely the same results as the classical spherical distribution.



Our first experiment concerns the measurement of the velocity of light over a closed path; in the path we shall insert refractive media in order to prove that the full range of velocities may be applied.

Fig. 2 shows a closed path $L^1, L^2, L^3, \dots, L^n$. From a point, S , on the path, we send a light signal in a clockwise direction along the path, and measure at S the time taken for the signal to cover the entire path once. According to the classical point of view, this time is :

$$t = \sum \frac{L}{v}, \quad \dots \dots \dots (5)$$

where the v 's are the respective velocities along the L 's; some of the v 's may be equal to c (3×10^{10} cm./sec.), while others are lower velocities through the refractive media, R_1 and R_2 .

The ellipsoidal distribution gives :

$$t = \sum \frac{L}{V} = \sum \frac{L}{v} \left(1 - \frac{v}{c} e \cos w \right) = \sum \frac{L}{v} - \frac{e}{c} \sum L \cos w, \quad \dots \dots (6)$$

in which the last term is the constant e , multiplied by the projection of the entire closed path on the polar axis—which, of course, is zero. The time is therefore

$$t = \sum \frac{L}{v}, \quad (7)$$

that is, the same as for the spherical distribution.

Our second representative experiment may be interpreted as an interference measurement. The apparatus is the same as in our first experiment, but this time we send two signals from S, one in each direction along the closed path. One of the signals, say the one in the anti-clockwise direction, is sent at an interval t later than the other signal. The two signals will meet somewhere along the path, say at P. We shall now show that the position of this point P is independent of the eccentricity and axis of the ellipsoidal distribution of velocities.

According to the classical point of view, we have the equation :

$$\sum \frac{L}{v} = t + \sum \frac{L}{v}, \quad (8)$$

which means that the two signals take equal times to arrive at their meeting point, if the time in each case is measured from the starting time of the first signal. From this equation we can find the position of the point P.

In an ellipsoidal distribution of velocities, the equation takes the form :

$$\sum \frac{L}{v} \left(1 - \frac{v}{c} e \cos w \right) = t + \sum \frac{L}{v} \left(1 - \frac{v}{c} e \cos w \right), \quad . . . (9)$$

which may be written :

$$\sum \frac{L}{v} - \frac{e}{c} \sum L \cos w = t + \sum \frac{L}{v} - \frac{e}{c} \sum L \cos w. \quad . . . (10)$$

In this expression the terms written last on each side of the sign of equation are equal. This is because they both are products of the same factors e/c , and the projection of the path—clockwise or anti-clockwise—from S to P on the polar axis. Equation (10) is thereby reduced to (8), which shows that the position of the point P is in both cases found from the same equation, and is therefore unaffected by the ellipsoidal distribution of velocities.

Another manner of carrying out the experiment is to send off the two signals simultaneously from S, and measure the time-lapse between their arrivals at point P; again we find that the result of our measurement is independent of the velocity distribution.

This brings us to Roemer's measurement.

Fig. 3 shows the distances involved in Roemer's calculation. E is the earth, here stationary because we, the observers, are standing on it. J_1 is an observed position of Jupiter where one of its satellites is in an easily observable position relative to Jupiter. J_2 is another position of

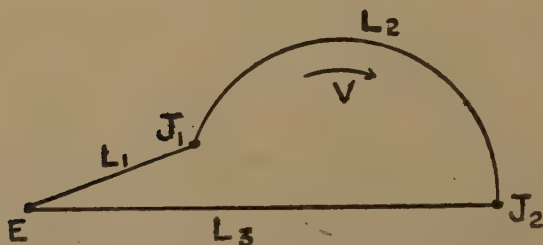
Jupiter further away from the earth, where the same satellite may be observed in the same relative position. The distance between J_1 and J_2 is usually chosen near its maximum value, *i. e.* near the diameter of the earth's orbit. The time measurement used by Roemer is the time difference between the arrivals at the earth, E , of the light from Jupiter and its satellite at J_1 , and that starting later from J_2 . Our task is to prove that this time difference is unaffected by the ellipsoidal distribution of velocities.

The time difference in a spherical distribution of velocities is expressed by the equation :

$$t = \frac{L_2}{v} + \frac{L_3}{c} - \frac{L_1}{c}, \quad (11)$$

in which the L 's may be found in fig. 3 ; v is the velocity of Jupiter in its apparent orbit. This ignores the earth's acceleration, the influence of which on the result is, however, so small that we can neglect it in the present treatment. In the proof given below, we assume that the eccen-

Fig. 3.



tricity of the ellipsoidal distribution of velocities is so great that the change of motion of the earth can be regarded as being of secondary importance. This in no way reduces the value of the proof, as it is, of course, at the very great eccentricities that we should expect the difference between spherical and ellipsoidal distribution to appear. The point of major importance in the present connection is that spherical and ellipsoidal distributions lead to the same result.

By introducing the ellipsoidal distribution of velocities, we get :

$$t = \int_0^{L_1} d\frac{L}{v} \left(1 - \frac{v}{c} e \cos w\right) + \frac{L_3}{c} (1 - e \cos w_3) - \frac{L_1}{c} (1 - e \cos w_1), \quad (12)$$

which may be written

$$t = \int_0^{L_1} d\frac{L}{v} - \frac{e}{c} \int_0^{L_1} dL \cos w + \frac{L_3}{c} - \frac{e}{c} L_3 \cos w_3 - \frac{L_1}{c} + \frac{e}{c} L_1 \cos w_1, \quad (13)$$

$$\text{or} \quad t = \frac{L_2}{v} + \frac{L_3}{c} - \frac{L_1}{c} - \frac{e}{c} \int_0^{L_1} dL \cos w - \frac{e}{c} L_3 \cos w_3 + \frac{e}{c} L_1 \cos w_1, \quad . \quad (14)$$

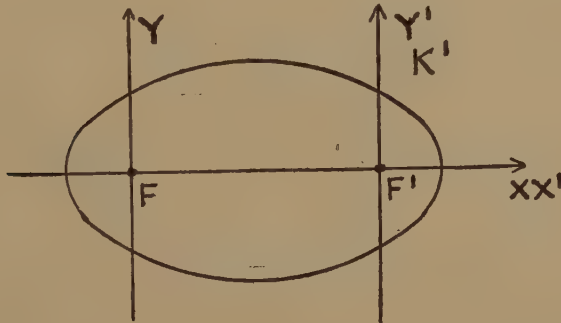
in which all the cosine terms are projections of the respective parts of the closed figure L_1, L_2, L_3 ; their signs are such that they cancel one another and leave an equation identical with (11). The time difference

in question is, therefore, independent of the ellipsoidal distribution of velocities.

A simple verbal explanation of this constancy is as follows:—If we increase the velocity of light towards the earth—in the direction from J_2 to E—we must decrease the velocity of light (and therefore smaller velocities also, as already explained) in the opposite direction in order to keep the there-and-back velocity constant. The result is, that Jupiter will take a longer time for its motion from J_1 to J_2 , and will, as it were, be late at J_2 . The light from J_2 to the earth, E, will, however, make up for the time lost by its relatively speedy journey towards the earth, thus making its actual velocity unknowable.

But, one may object, if Jupiter is late at J_2 , we should be able to verify it by observation from the earth. The truth is that we cannot. The light we would use for our observation is, as we have just seen, speeded up so that it arrives at the earth at a time, and from a direction, which makes it look as if Jupiter were exactly on time.

Fig. 4.



Roemer's measurements and his method of calculation therefore do not give the one-way velocity of light; they give only the there-and-back velocity. This, although no more conclusive than other proofs by examples, should strengthen our belief in the postulate making the velocity of light an unknowable, sufficiently to make us attempt a reformulation of the special theory of relativity, usually based on the belief that the velocity of light is measurable, and constant, in all directions.

Let us assume that an ellipsoidal distribution of velocities exists in an inertial frame of reference, K, represented by the Cartesian co-ordinate system x, y in fig. 4, the x -axis of which is the polar axis of the distribution.

At the time $t = 0$, a light signal is emitted in all directions from the point of origin of the co-ordinate system. The front of this signal will spread as the surface of a growing ellipsoid, so that at the time t this front in the x, y plane is the ellipse shown. The focal point F of the ellipse coincides with the point of origin of the frame K.

But it is not only relative to the focal point F that the distribution is

ellipsoidal in accordance with (1); the other focal point F' is, of course, an equally valid pole. If we therefore introduce another frame of reference, K' , moving with the same velocity as the focal point F' of the growing ellipsoid, we find that this frame K' is entirely equivalent to the frame K with respect to the measurable characteristics of the velocity of light. We have here what may be termed symmetrical relativity. Units of length and time are the same for both frames, simply because the two frames and their velocity distributions are identical in structure.

The velocity of the frame K' is found in the following manner.

The focal distance of an ellipsoid as described in (1) is obviously the difference between the two radii along the polar axis. We therefore have :

$$\text{focal distance} = \frac{c}{1-e} - \frac{c}{1+e} = \frac{2ce}{1-e^2}. \quad (15)$$

This distance is equal to the vector representing the velocity of the frame K' . If that velocity in a spherical distribution of velocities is u , then it is, in the ellipsoidal distribution of the frame K , expressed by :

$$U = \frac{u}{1 - \frac{u}{c} e \cos w} = \frac{u}{1 - \frac{u}{c} e}, \quad (16)$$

because $w = 0$.

From (15) and (16) we get :

$$\frac{2ce}{1-e^2} = \frac{u}{1 - \frac{u}{c} e}, \quad (17)$$

which gives

$$e = \frac{c}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right). \quad (18)$$

(Only the negative value of the root is used, as $0 < e < 1$ for an ellipsoid.)

By substituting e from (18) in (16) we find

$$U = \frac{u}{\sqrt{1 - \frac{u^2}{c^2}}}, \quad (19)$$

indicating that the distortion which increases the velocity, u , between

the two frames of reference to $\frac{u}{\sqrt{1 - \frac{u^2}{c^2}}}$ distorts at the same time the

distribution of the velocities of light exactly so that any there-and-back measurement of the velocity of light will give the same result in either frame.

Students of special relativity will note the appearance of $\sqrt{1 - \frac{u^2}{c^2}}$.

As we go on, we shall see that it will appear in all measurable values in which it has been found experimentally.

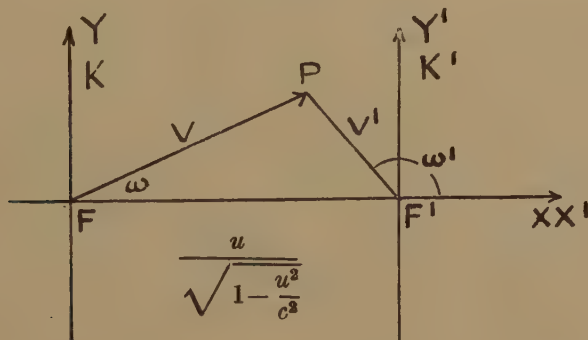
We can now find the final expression for V given in (4), by using (18). The result is :

$$V = \frac{v}{1 - \frac{v}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right) \cos w} \quad (20)$$

This allows us to work out an expression for the composition of velocities. Fig. 5 shows two frames of reference, K and K' , in relative motion with the velocity $\frac{u}{\sqrt{1 - \frac{u^2}{c^2}}}$, corresponding to the velocity u in

a spherical distribution of velocities in either of the frames. A point P moves relatively to the frame K , with a velocity V making the angle w

Fig. 5.



with the x -axis of the frames. In the frame K' , the motion of the point P is measured by the velocity having the polar components V' and w' .

By projection on the x -axis we find :

$$\frac{V \cos w}{1 - \frac{V}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right) \cos w} = \frac{u}{\sqrt{1 - \frac{u^2}{c^2}}} + \frac{V' \cos w'}{1 + \frac{V'}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right) \cos w'} \quad (21)$$

which reduces to

$$V \cos w = \frac{u + V' \cos w'}{1 + \frac{u V' \cos w'}{c^2}} \quad (22)$$

or in the more familiar form

$$V_x = \frac{u + V_x'}{1 + \frac{u V_x'}{c^2}} \quad (23)$$

By projection on the y -axis, we have :

$$\frac{V \sin w}{1 - \frac{V \cos w}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}}\right)} = \frac{V' \sin w'}{1 + \frac{V' \cos w'}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}}\right)}, \quad (24)$$

which gives

$$V_y = \frac{V'_y \sqrt{1 - \frac{u^2}{c^2}}}{1 + \frac{u V_x}{c^2}} \quad \dots \dots \dots (25)$$

This, of course (see also (23)), is the well-known expression from the special theory of relativity.

All the problems that are solved in the special theory of relativity by application of the formulæ for composition of velocities only are, therefore, solved in the same manner in the new transformation system. For example, aberration is treated in exactly the same manner in both theories, namely, by postulating that the velocity of light is not influenced in a measurable manner by the motion of the emitting body, and applying the formulæ for composition of velocities, which then leads directly to the result.

Another important problem is to decide how the unit of mass is transformed from one frame of reference to another. Here, again, we follow the course taken in the special theory of relativity. We postulate the constancy of kinetic energy and momentum, and carry out a simple imaginary collision experiment with two masses. We apply the formulæ for composition of velocities, and find that momentum has the magnitude

$$\frac{mu}{\sqrt{1 - \frac{u^2}{c^2}}} \text{ instead of } u \times m \text{ as presumed in Newtonian mechanics.}$$

This result is interpreted in the special theory of relativity as indicating that the mass is increasing with velocity. But if we assume the ellipsoidal distribution of velocities, we may keep mass as an invariant, because the

increased velocity $\frac{u}{\sqrt{1 - \frac{u^2}{c^2}}}$ fully accounts for the increase in momentum.

Other problems, although treated differently in the two theories, give the same results in both. A good example of this is the Doppler effect.

Let us assume that a point F' (see fig. 6) is moving relatively to another point F with the velocity $\frac{u}{\sqrt{1 - \frac{u^2}{c^2}}}$ in an ellipsoidal distribution of velo-

cities, and that at time intervals t , signals are emitted from F' . For the sake of simplicity, we shall let one signal be emitted at the instant of time when F' coincides with F . The question is now, how long time does it take at F , before the next signal from F' arrives ?

The waiting time at F is the sum of two times. One is the time t , in which F' moves $t \times \frac{u}{\sqrt{1-\frac{u^2}{c^2}}}$ to the point from which it emits the signal

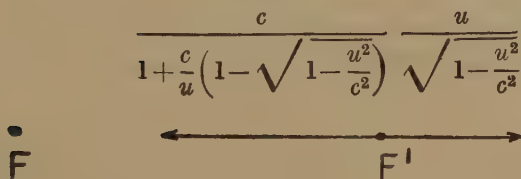
in question, which then moves towards F with the velocity

$$\frac{c}{1 + \frac{c}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right)}$$

The other, the time it takes for the signal to travel back to F, is therefore

$$\frac{tu}{c \sqrt{1 - \frac{u^2}{c^2}}} \left(1 + \frac{c}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right) \right), \quad \dots \quad (26)$$

Fig. 6.



and the sum is

$$T = t + \frac{tu}{c \sqrt{1 - \frac{u^2}{c^2}}} \left(1 + \frac{c}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right) \right), \quad \dots \quad (27)$$

or

$$T = t \frac{\sqrt{1 + \frac{u}{c}}}{\sqrt{1 - \frac{u}{c}}}, \quad \dots \quad (28)$$

which is the result known from the special theory of relativity.

We have seen, in these examples, how different views may lead to the same result; and it must be noted that the results obtained are experimentally verifiable. There is, however, another type of problem, in which the result, or at least part of it, appears as an unverifiable construct, no less important than its verifiable part, because it concerns the relation of relativity to other physical theories.

A good example appears in connection with the notion of simultaneity. In the new theory, this takes on an unexpectedly simple form, which may be shown by comparing the applications of the two theories to the same experiment.

The experiment is the simple one of two rods of equal rest-length passing one another, as shown in figs. 7 and 8.

Fig. 7 shows the experiment, as we must regard it if we adopt the spherical distribution of velocities as postulated in the special theory of relativity. One rod, $A' A''$, is at rest, while the other one, $B' B''$, is passing it with the velocity u . The moving rod is, of course, contracted to $\sqrt{1 - \frac{u^2}{c^2}}$ of its rest-length. We arrange for two light signals to be

sent, one from each end of the resting rod towards its centre A . One signal is sent from A'' at the moment when B'' passes, the other one is sent from A' when B' passes. The time lapse between the arrivals of the two signals at A is, of course, the time it takes for B' to travel from

Fig. 7.

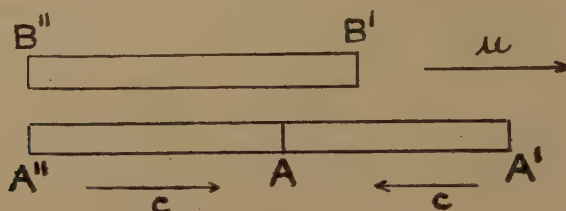
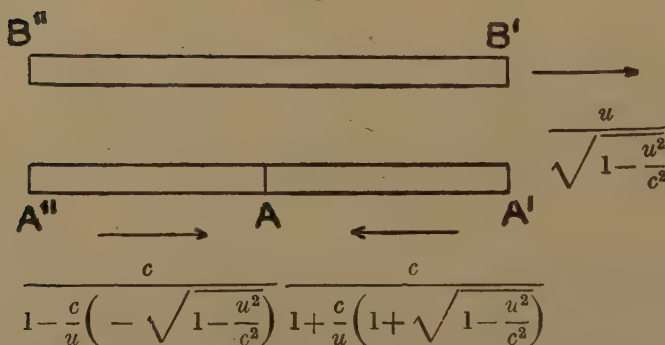


Fig. 8.



the position shown in fig. 7 to the position where B' is opposite A' . This time lapse is :

$$T = \frac{l \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right)}{u}, \dots \dots \dots (29)$$

where l is the rest-length of each rod.

It is evident that, from this point of view, we may in no circumstances regard the sending of the two signals as taking place simultaneously.

In fig. 8, we see how the same experiment must be understood from the new point of view. The velocity of the rod is now $\frac{u}{\sqrt{1 - \frac{u^2}{c^2}}}$, and there

is no contraction of length. The ends of the two rods cover one another simultaneously. The time lapse between the arrivals of the two signals at A is found simply by calculating the difference between the times it takes for the two signals to cover their respective paths—remembering that the signals now travel with different velocities. The difference is :

$$T = \frac{l}{2c} \left(1 + \frac{c}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right) \right) - \frac{l}{2c} \left(1 - \frac{c}{u} \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right) \right), \quad (30)$$

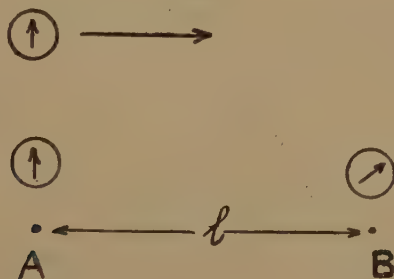
or

$$T = \frac{l \left(1 - \sqrt{1 - \frac{u^2}{c^2}} \right)}{u}, \quad \dots \dots \dots (31)$$

that is, the same result as we found by using the orthodox theory. The verifiable result—the time-lapse—is thus the same in both theories, whereas the unverifiable notion of simultaneity is entirely different.

In order to be quite clear about the difference in our conceptions of simultaneity in the two theories, let us see how we interpret the time lapse in the new theory.

Fig. 9.



We receive two signals, indicating the occurrence of two events. One signal arrives at an interval t later than the other. We know, however, that the distances and velocities involved are such that the time of travel of the last signal will be longer than that of the first by this same interval, t . We conclude, therefore, that the two events must have been simultaneous in the frame of reference at rest relatively to the rod A' A''. Owing to the complete symmetry of the two frames of reference, events simultaneous in one of them must also be simultaneous in the other.

It is outside the scope of this paper fully to discuss the influence of the above result upon the position of special relativity in relation to other parts of modern physics, but the reader will understand that it is on this point that the new theory may prove itself more applicable than the classical theory.

Another well-known relativistic effect is the slowing down of the rate of a moving clock. The measurable effect is, in this case also, the same, for both theories, but our conception of the effect is different.

Consider the system shown in fig. 9. Here, two points in space, A

and B, have each a clock ; a third clock is moving in the direction from A to B with a constant velocity.

The A, B system is at rest in a frame of reference, K ; the moving clock is at rest in a frame K'. We know from our consideration of the symmetry of the two frames that the time units in both are equal. This would normally mean that a moving clock would always read the same as any resting clock it was passing—provided, of course, that we have not set the resting clocks “wrongly.” But that is exactly what we do when we introduce the ellipsoidal distribution of velocities : we adjust our clocks in space, in order to alter the numerical value of the velocity of light—and thereby all other velocities. The clock B will, if the

velocity of the moving clock is $\frac{u}{\sqrt{1-\frac{u^2}{c^2}}}$ in the ellipsoidal distribution of

velocities, be advanced $\frac{l}{u} - \frac{l}{u} \sqrt{1-\frac{u^2}{c^2}}$ over what it would read in a

spherical distribution of velocities. This means that the clock at rest at B will record a later time than the moving clock when it passes B. This is what we should expect from the classical point of view, where it is interpreted as indicating that the resting clock is “right,” while the moving clock is “slow,” *i. e.* working at a slower rate. From the new point of view it is the moving clock that is “right,” and the resting clock that is—as we know—advanced, not fast in the sense of working faster, but simply by having its hands moved forward.

We have seen how it is possible, by introducing the ellipsoidal distribution of velocities, to retain the units of length, time and mass as invariants in the double system we have been considering. But how do the units fare in a frame of reference other than the two forming the partial absoluteness ?

To answer this, let us see how we may measure motion in our double system of absoluteness. We choose, in our K' frame, a unit length, and measure its state of motion relative to the K frame by observing at one point in the K frame—with a single clock—the time it takes for this length to pass.

This single-clock measurement must, naturally, give the same result for all directions, irrespective of how we choose the distribution of velocities. This leads to a new kind of contraction of length due to velocity, which is the correction necessary to make the single-clock measurement of motion give the same result for all directions. It takes the place of, and contains, the Fitzgerald contraction.

The value of the single-clock velocity of a unit of length in motion is, naturally, the value we use in the corresponding ellipsoidal distribution.

This velocity is $\frac{v}{\sqrt{1-\frac{v^2}{c^2}}}$, corresponding to the velocity v in a spherical

distribution. If, now, we want to find the contraction of a unit length travelling in a direction given by the angle w in another ellipsoidal distribution of velocities, we must arrange for the velocity in this direction, in all cases, to be compensated by the contraction. This is done by making it the ratio between the two velocities involved. We have, therefore, by using (20),

$$l_{u,w,v}=l_0 \frac{\frac{v}{1-\frac{v}{u}\left(1-\sqrt{1-\frac{u^2}{c^2}}\right)\cos w}}{\sqrt{1-\frac{v^2}{c^2}}}, \quad \dots \quad (32)$$

which may be written

$$l_{u,w,v}=l_0 \frac{\sqrt{1-\frac{v^2}{c^2}}}{1-\frac{v}{u}\left(1-\sqrt{1-\frac{u^2}{c^2}}\right)\cos w}. \quad \dots \quad (33)$$

By a similar manner of reasoning about a single clock in the K' frame passing a unit length in the K frame, we find the slowing down of the rate of time outside our frames of absoluteness. This is given by

$$t_{u,w,v}=t_0 \frac{\sqrt{1-\frac{v^2}{c^2}}}{1+\frac{v}{u}\left(1-\sqrt{1-\frac{u^2}{c^2}}\right)\cos w}. \quad \dots \quad (34)$$

In our absolute frames, these corrections become unity, because $w = 0$ and $u = v$.

If $w = \frac{\pi}{2}$, we have

$$l_v=l_0\sqrt{1-\frac{v^2}{c^2}} \quad \text{and} \quad t_v=t_0\sqrt{1-\frac{v^2}{c^2}}, \quad \dots \quad (35)$$

indicating that classical special relativity exists in the y, z directions. This is, of course, to be expected, as the velocity of light in any plane parallel to the y, z plane is c (see fig. 1).

These corrections do not, however, detract from the simplicity of the partial absoluteness of the two frames; they constitute a background, which in many cases need not be considered at all. They show, on the other hand, the simple connection between the mathematics of the two theories; the new mathematics may, in fact, be derived from the Lorentz transformation formulæ.

It is tempting to rationalize all these results in a complete reformulation of the special theory of relativity, but as the writer is at present unable to find sufficient time for such work, he has decided to prepare these introductory notes in a less complete form.

Conclusion.

By choosing special values for the velocity of light in different directions in space, it is possible to make the mechanics of any two inertial frames of reference absolute—time rate, length, mass and simultaneity are absolute in the two chosen frames. The most important result is the absoluteness of simultaneity, which should facilitate the smooth blending of special relativity with quantum and wave mechanics. These changes in the expression of the special theory of relativity are brought about solely by postulating an unknowable variation in the one-way velocity of light.

Einstein's postulate of the constancy of the velocity of light *in vacuo* should be modified to cover velocities measured over closed paths only, and should not be applied to velocity vectors. With this modification an unnecessary and impeding rigidity may be removed from the meaning of the special theory of relativity. The measurable results remain unaltered.

XIV. *New Contributions to Interferometry. Part II—New Interference Phenomena with Newton's Rings.*

By S. TOLANSKY, Ph.D., D.I.C., The University, Manchester*.

[Received November 1, 1943.]

[Plates I.–III.]

SUMMARY.

New interference phenomena are observed with Newton's rings when the lens and glass plate are coated with high reflecting thin silver films. Very sharp rings are formed in both transmission and reflection, the intensity distribution being analogous to the Fabry-Perot interferometer fringes. Very great precision in measurement can be achieved. In addition, surface errors of less than one-hundredth of a wave-length can easily be detected on lenses. At incidences other than normal, the fringes, which are no longer localized in the interference film, become double. The separation of the doublet increases with the angle of incidence. The doubling is shown to be due to differential phase change at reflection of the light polarized in and perpendicular to the plane of incidence. The components of the doublet are plane polarized at right-angles to each other. With increasing incidence the parallel magnetic vector component becomes steadily weaker. This is shown to be due to a preferential absorption effect. A single thin silver film at non-normal incidence is shown to transmit partially polarized light, agreeing

* Communicated by the Author,

with the above observations. The behaviour of both the reflected and the transmitted systems can be accounted for in detail.

An analogous doubling is shown to exist with thin film wedge fringes.

A similar doubling is recorded for Fabry-Perot fringes when viewed at increasing incidences.

The new fringes have a number of possible applications. Metrological work employing Newton's rings can be considerably improved in precision. A new method for testing lens surfaces with high precision is available. The differential phase change at reflection can be objectively determined, for the first time without measurements of elliptical polarization effects.

Introduction.

THE study of Newton's rings, formed by illuminating a lens placed upon a glass plate, has produced an extensive literature of which a comprehensive survey has been given by Feussner ⁽¹⁾ and by Maclaurin ⁽²⁾. With but few exceptions the simple reflected system has been studied, although some observers have replaced the plane glass plate by a solid metallic mirror, still making observations upon the reflected system.

It occurred to the author, some eight years ago, that Newton's rings could be studied in transmission with considerable advantages if high reflecting coefficient thin transparent silver films (of the type used for the Fabry-Perot interferometer) were deposited on the contacting faces of the lens and glass plate. Such an arrangement was tried then, and gave striking results. It is the object of this paper to describe the modifications produced by silvering the two surfaces and to report or unusual new interference phenomena recently observed with such apparatus. These phenomena afford a powerful new method for studying some of the optical properties of thin metallic films. The general principles will be described here, and in the succeeding paper (henceforth entitled Part III.) quantitative measurements for a particular silver film will be discussed in detail.

Newton's Rings as Modified by Multiple Reflections.

The essentially important experimental procedure needed for the study of the modifications introduced into Newton's rings by multiple beams is the employment of silverings with *high reflecting coefficients*. *Half silverings are totally inadequate*. The requisite technique for the production of high quality mirrors had already been developed some years earlier for use with the Fabry-Perot interferometer applied to hyperfine structure studies.

The mirrors, of silver, some 5×10^{-6} cm. thick, were deposited upon the lens and glass plate simultaneously by evaporation *in vacuo*, the usual precautions adopted for Fabry-Perot interferometry being adhered to. The reflecting coefficient was of the order of some 0.85 or more for the green mercury line.

The deposition of such silver films upon the optical components leads to remarkable changes in both the transmitted and the reflected interference fringe systems. The discussion which follows will be largely concerned with the transmitted system, which is easy to study experimentally. The reflected system has only been examined in a qualitative manner, the quantitative examination being made with the transmitted fringes.

With unsilvered glass components the visibility of the transmitted system is, of course, very poor, since the two interfering beams have widely differing intensities. With silvering, multiple reflections take place. The lateral displacement of successive beams which takes place when a lens of long focal length is used is a small fraction of the linear distance between adjacent rings. Effectively, therefore, all the multiply-reflected beams contributing to a given fringe can be considered to combine locally in one circle. This is therefore analogous to the condition in the Fabry-Perot interferometer, or to be more precise, the interference takes on a similar form to that in Fabry's modification of Fizeau's localized wedge fringes, but are circular instead of straight. The rings, therefore, have the characteristic highly sharpened intensity distribution of a Fabry-Perot interferometer with good silverings.

The fringe visibility is therefore extremely high, and further, the intensity of the transmitted system exceeds the intensity of the reflected system given by unsilvered surfaces. With a single-line source brilliant extremely fine rings on a broad dark background can be seen localized in the interference film either by the naked eye or with a microscope. Their *appearance* is similar to that of Fabry-Perot rings, and if a mercury arc source is used, unfiltered, the separate lines form differently coloured rings of different radii, producing a colourful pattern.

It has been known since the time of Newton that the diameters of the rings increase with θ , the angle of incidence of the light, being proportional to $\sec \theta$. An incidence of 10° leads to a 1.5 per cent. increase in diameter. So sharp are the silvered rings that the definition is affected by the size of the source. With the extended sources usually employed for Newton's rings a range of angles of incidence exceeding 10° can easily fall on the interference film. With unsilvered reflected fringes, changes due to this are hardly noticeable. With the transmitted silvered fringes a clear, noticeable broadening results, and in particular if the light is not incident normally, a considerable effect due to the extended source appears. Thus, high grade definition is only obtained if the incident light is restricted to an approximately single angle of incidence, and this is simplest achieved by employing a small source (say not exceeding 5 mm. in diameter, preferably less, at non-normal incidences) at the focus of a lens, projecting the approximately parallel beam onto the interference film. This parallel beam collimation is a matter of some importance when the light is incident at angles exceeding 45° .

Under suitable conditions a reflected system, complementary to the transmitted system, can also be seen. This consists of extremely fine

dark rings on a broad bright background, almost like fine absorption lines. The location of the dark reflected system is identical with that of the bright transmitted fringes. The visibility depends upon special factors which will be discussed later. With an unfiltered line source such as a mercury arc, the separately coloured rings can easily be seen in transmission, but the reflected system is only clearly defined with monochromatic radiation, for obvious reasons.

The Transmitted System. Normal Incidence.

Plate I. is a photograph of the transmitted system given by a glass plate and a lens with radius of curvature of the contacting surface equal to some 18.5 metres. The light used is the green mercury line filtered from a vacuum arc. The incidence is normal. The characteristic Fabry-Perot distribution is obvious. So inherently sharp are the rings that all surface defects on the lens and plate are rendered visible. An ordinary piece of good quality plate glass was used, and the small scale defects on this are clearly shown by the uneven character of the rings. It should be remembered that the distance between adjacent rings means an alteration of half a wave-length in the thickness of the air film. It is clear that defects of the order of one two-hundredth of a wave-length are rendered visible by this procedure.

At local regions, where the glass components are sensibly free from defects, the rings are remarkably sharp. They are, in fact, sharper than what one obtains *in practice* with a Fabry-Perot interferometer with the same silverings. The reason is obvious. In the Newton's ring the fringes are localized and the multiple reflections occur over a very small area of the glass. In the Fabry-Perot case, the light is integrated over a relatively large glass area, all light incident at one angle, over *all* the interferometer, contributing to the fringe. In practice, therefore, the Fabry-Perot rings never attain theoretical sharpness because of inevitable surface defects.

There is no doubt at all that given glass components of high optical quality, very high precision in the measurement of ring diameters would result. The *differences in ring diameters* can be determined to within 1 in 10,000, the actual diameters with higher precision still, by at least a factor of 10. This precision is not possible with crude glass plates since the fringes follow the contours of the irregularities and (on a small scale of course) actually zig-zag around a mean circle.

Effects arising with Incidence other than Normal.

Striking new interference effects have been found by altering the angle of incidence of the parallel beam falling upon the silvered plate-lens combination. This is achieved by tilting the latter, the glass plate being towards the source to avoid changes in incidence which would arise with light incident on the curved lens surface. Tilting should lead to the appearance of elliptical fringes since the rings are now viewed off axis.

This is indeed the case. Furthermore, in accordance with simple geometrical construction, a change in the location of the fringes should take place. Feussner⁽⁵⁾ has shown that, with wedge fringes, if the wedge apex points *away* from the observer the transmitted fringe system is virtual, the fringes being located behind the wedge. If the wedge apex points *to* the observer real fringes are formed in front of the wedge. Observation shows that the transmitted Newton's rings at non-normal incidence lie on regular curves, one half before and the other half behind the interference film. The question of the exact location will be discussed in detail later. As a result of the peculiar focal position in space, a photographic plate set normal to the light beam records only a few fringes simultaneously in focus, the true focal plane lying often at a very steep angle when the fringes are projected with a lens.

As the angle of incidence of the light is gradually increased a new effect begins to make its appearance and can be just detected first at an incidence of 20° . *The fringes begin to double.* At 30° the doubling is quite clearly marked and as the incidence increases the separation between the two components increases steadily. One component appears to detach itself and move across towards the next higher order. The march of the outer component with increasing incidence is shown clearly in Plate II., which illustrates the appearance at normal incidence, and at 30° , 35° , 40° , 45° 80° .

Attention may be drawn to three special characteristics. As the incidence increases

- (a) The doublet separation grows regularly.
- (b) When first resolved, both components have the same intensities, but as the outer higher order fringe moves away it becomes progressively weaker, finally vanishing as it approaches the next order.
- (c) As the incidence increases, the outer component becomes progressively sharper and sharper, being extremely narrow at the higher angles of incidence. (This is to be observed best in the small region, where the focus is correct. Visually it is strikingly obvious.)

The sharpness of the complete system of the outer fringes at incidences exceeding 60° is unique. It is fairly certain that no other *complete* system, equally sharp, can be obtained by other means (the double *Fabry-Perot* fringes to be described later are equally sharp, for the same cause, and can be considered optically equivalent for this purpose of comparison). Related fringes will be considered later.

Although at normal incidence the fringes are quite sharp and narrow, the sharpness of *both* members of the doublet increases with incidence. The stronger fringes only sharpen up slowly whilst the outer weaker systems sharpen up at a very rapid rate. This striking difference in fringe width can only be interpreted as meaning that the reflecting coefficients for the two sets are increasing with incidence but at very different rates. The fringe width is extremely sensitive to the value of the reflecting coefficient, as is well known from the corresponding *Fabry-Perot* interferometer case.

This observation at once suggested that the two sets of rings consist respectively of light polarized at right-angles and parallel to the plane of incidence, the outer sharper rings being formed by the beam, which has the magnetic vector parallel to the plane of incidence. This was proved to be correct by the introduction of a Nicol prism or polaroid disk into the incident beam. The effect is shown clearly in Plate III.

Plate III. shows the appearance of the fringes with angles of incidence 55° and 60° . A triple shutter was placed over the photographic plate and the sections A, B, C exposed in turn. B shows the fringes with no Nicol or polaroid in the field. At A the Nicol or polaroid is set to pass the vibrations which are polarized in the plane of incidence and at C polarized perpendicular to the plane of incidence. If a Nicol is used, care must be taken not to alter the angle of incidence because of the deviation of the prism. A polaroid disc obviates this difficulty. In Plate III. (a), incidence 60° , a Nicol prism was used, and a very slight displacement can be detected, produced by the rotation of the prism, leading to a slight alteration in incidence. In Plate III. (b), incidence 55° , a polaroid disc was employed. There is no fringe displacement. The exposures have been adjusted to bring out the effect of the polarizer. In both cases the weaker parallel vector component, at A, has been given twice the exposure of the stronger at C. The correct intensity ratio is seen at B.

The Origin of the Doubling of the Rings.

The doubling of the fringes arises because of the phase change taking place at reflection at a metallic surface. It is well known from classical electromagnetic theory that in the case of non-normal incidence the phase change is different in the case of parallel and perpendicularly polarized light. At normal incidence the differential phase change is zero. As the incidence increases the differential phase change grows. It is quite clear that this differential phase change leads to the observed doubling, since it effectively alters the optical path between the metal surfaces.

In general the lens, of radius R, is not in perfect contact with the glass plate. Let the two glass components be separated by a distance τ in a medium of unit refractive index. Let ϵ be the effective path change at a *single* reflection. Then the path difference between the succeeding beams, which add up to form the infinite series due to multiple reflections, is

$$2t \cos \theta + 2\tau \cos \theta + 2\epsilon = n\lambda,$$

in which θ is the angle of incidence. If ρ_n is the radius of the n th ring then

$$\rho_n^2 = \frac{R}{\cos \theta} (n\lambda - 2\tau \cos \theta - 2\epsilon),$$

from which it follows that an increase in the ring diameter arises from a negative phase change, *i. e.* one which effectively reduces the optical path.

If we differentiate this we obtain the differential phase change, for we have $d\epsilon = dn \cdot \lambda/2$, thus the differential phase change, $d(\epsilon/\lambda)$, is $dn/2$, which is half of the observed fraction of an order separation between the two fringe components. This is thus independent of the ring order number, the radius of curvature, and does not explicitly involve θ , although it is a function of the angle of incidence.

The differential phase change at reflection can thus for the first time be directly *objectively* determined from the fringe displacement. In the former, now classical, determinations of this quantity, the phase shift was not directly observable but was derived from measurements of the degree of elliptical polarization of the light reflected from a metallic film.

The direct determination can only be carried out because of the extreme sharpness of the rings, which owe their fineness to the very metallic films introducing the phase change. The quantitative evaluation of $d(\epsilon/\lambda)$ for a specific silver film is described in Part III.

The Location of the Fringes.

The location of the transmitted fringes at non-normal incidence presents interesting features. To a first approximation the fringes can be considered as wedge fringes, since each fringe is localized and arises from multiple reflections over a small area. The effective wedge-angle is not constant but increases progressively with fringe order number. It can be considered that at any point the interference is that of a wedge, the angle of which is the angle made by the tangent to the lens at the point in question, with the plane glass surface.

Considering the Newton's rings (that is, considering a single ring only) as arising from a double wedge, it is clear from the construction shown in fig. 1 that at non-normal incidence the wedge AOB leads to a fringe formed by multiple reflections at the point X, whilst the wedge COD leads to a fringe located at Y. Hence one-half of the ring system will be in front of and the other half behind the interference film.

The fringes corresponding to the next order will lie on the same sides as above, but as the equivalent wedge-angle has changed, the distance of the fringe from the interference film changes. The following approximate theory of the fringe location is in qualitative agreement with the observed positions.

Feussner has shown that the fringes formed by multiple reflection of light incident at an angle θ on to a wedge of thickness d and wedge-angle ϕ , with refractive index μ , appear at a distance D from the plate given by

$$D = \frac{\sin \theta \cdot \cos^2 \theta}{\mu^2 - \sin^2 \theta} \cdot \frac{d}{\phi}.$$

For an air film this reduces to $D = \sin \theta \cdot d/\phi$. It is legitimate to apply this to the Newton's ring fringes, using for each fringe the film thickness as d , and the angle between the plane and the tangent as ϕ .

For the n th ring the wedge thickness is $t + \tau$ (former notation), where $n\lambda = 2(t + \tau) \cos \theta$. The angle ϕ between the tangent and the plane is

given by $\sin \phi \doteq \phi = \rho_n/R = 2t/\rho_n$, where ρ_n is the radius of the n th ring. The distance D_n of the n th ring from the interference film is thus

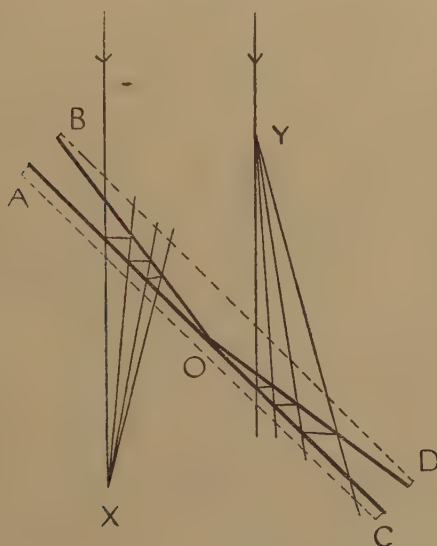
$$D_n = \sin \theta \cdot \frac{t + \tau}{2t} \cdot \rho_n.$$

Since the fringes are viewed in the direction of incidence, they appear as ellipses, and the measured radius of the smaller axis r_n is $\rho_n \cos \theta$. Thus we have

$$D_n = \frac{r_n}{2} \cdot \left(1 + \frac{\tau}{t}\right) \cdot \tan \theta.$$

The argument applies to both halves of the pattern, which will be on opposite sides of the interference film. (τ can be taken for this analysis to include the phase change at reflection.)

Fig. 1.



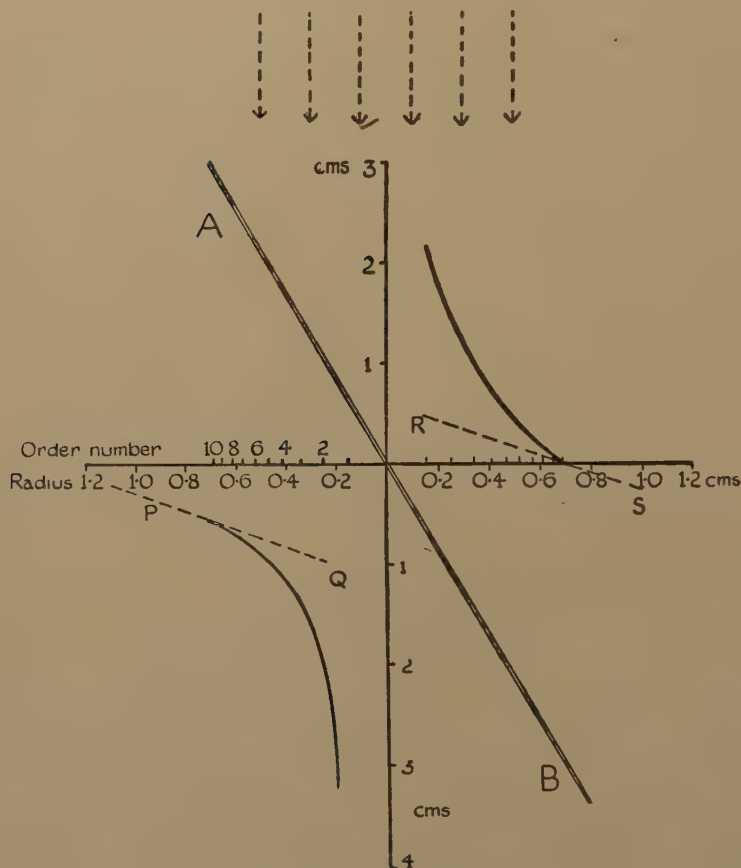
Formation of double wedge fringes.

Examining this in detail the following conclusions can be drawn. At the centre of the system (where $t=0$) the fringes will be at infinity. As t increases, that is, as the ring order number increases, the distance D_n diminishes with great rapidity and ultimately the fringes lie on a line with slope $(\tan \theta)/2$ when the distance D_n is plotted against the observed ring radius r_n .

The qualitative agreement between this much simplified theory and the observed fringe location is shown in fig. 2. The double line AB shows the plane taken up by the interference film. The incident light (angle 60°) is shown coming down from above. With a specific case in which the radius of curvature of the lens $R=18.5$ metres, the location of the first 10 fringes is that shown by the two curves in the first and third quadrants. It will be noticed that the curves come down apparently

from infinity at the centre of the system, one before and one behind the interference film, as predicted. At the first ring, the difference in location from both sides is still considerable, namely 5 centimetres. The lines PQ, RS, are drawn at angles with AB, the interference film equal to $(\tan 60^\circ)/2$, and it is clear that the fringes do ultimately lie approximately along these lines, as the simple theory requires.

Fig. 2.



Location of the fringes.

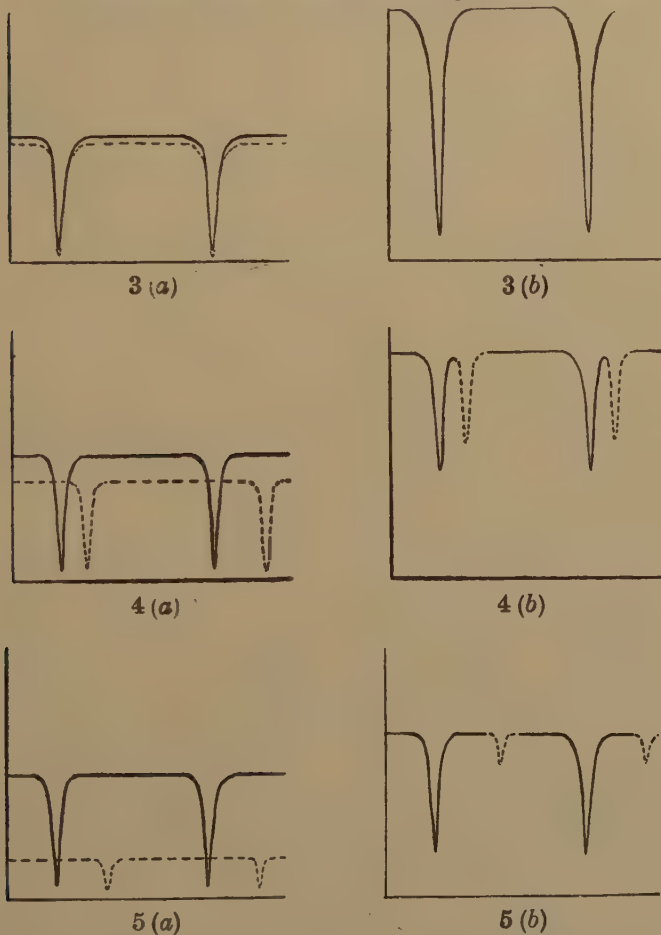
It is thus evident that the values of D_n depend upon θ in a manner which makes D_n grow very rapidly when θ increases beyond 60° . This makes the photography of the fringes at large values of θ a matter of difficulty. If the fringes are projected by a lens on to a photographic plate set normal to the axis only a few fringes are simultaneously in focus. If more are to be in focus then the plate must be tilted through large angles. It has been found convenient to set the plate normal to the axis, and in the reproductions in Plate II. and Plate III. only some three or four fringes, near the centre of the pattern, are in focus.

The Reflected Fringe System.

In addition to the transmitted system there exists a complementary reflected system, the form of which, in accordance with Hamy's⁽³⁾ early investigations on back reflected systems, depends upon the thickness of the silver film. With the high reflecting coefficient films employed here the reflected system consists of very fine dark rings upon a bright back-

Figs. 3 (a)–5 (b).

Reflected system with increasing incidence,



ground. This is complementary to the transmitted system as regards position and shape, but the overall intensities, as Hamy first pointed out, depend upon the film thickness.

With increasing incidence the fine dark reflected rings split up in exactly the same way as do the transmitted fringes, because of the differential phase change at reflection. As in the transmitted system, the outer rings become progressively sharper and weaker as the incidence increases, until they are submerged in the bright background.

If observations are begun at or near normal incidence, the *visibility* of the reflected fringes at first begins to diminish as the incidence is increased. When, however, a Nicol prism is introduced into the incident beam, either set to pass light polarized in or perpendicular to the plane of incidence, there results a striking change in fringe sharpness and visibility. It is clear from figs. 3, 4, 5, why this is so. At normal incidence the system due to the beam with the magnetic vector parallel to the plane of incidence (dotted) superposes on the perpendicularly polarized beam. For the sake of clarity the fringe patterns are *vertically* displaced slightly in fig. 3 *a*. The resultant is the sharp composite fringe shown in fig. 3 *b*. Increasing the incidence leads to the situation shown in fig. 4 *a*. It is clear that the fringes of each system are now superposed upon the strong background of the other, the resultant, in fig. 4 *b*, leading to a double set of fringes of considerably reduced visibility (this does *not* take place in the transmitted system). At higher incidence the appearance is that shown in fig. 5. It is important to note that when the Nicol is set to permit the weak fringes to pass, the *background* intensity still remains high at increasing incidence whilst the fringes become less and less visible.

The Intensity of the Variable Component.

A striking feature of the doubling of the fringes in both the transmitted and the reflected systems is the rapid falling off in the intensity of the fringes due to the beams polarized with the magnetic vector in the plane of incidence, as the angle of incidence is increased. No exact intensity determinations have been made, but visual estimates of the intensity variation are discussed in Part III. At angles above 80° the parallel vector fringes can barely be detected.

Since the intensities show the same trend in both the transmitted and the reflected systems, the effect can only be explained by postulating a rapid *relative* increase in *absorption* coefficient for the parallel magnetic vector. The increase in *reflecting* coefficient for the parallel vector does not, of course, account for the fall of intensity. For although less light is available for transmission, with a higher reflecting coefficient, yet it is well known that the Airy formula shows that the intensity of the maxima, if *absorption is disregarded*, is independent of the value of the reflecting coefficient.

The simultaneous disappearance of both transmitted and reflected fringe systems proves that the *absorption* of the multiply reflected beams is responsible for the intensity change. The multiply reflected transmitted beams each pass through two silver films, and if there is a differential absorption, one system is weakened. In reflection, the first strongly reflected beam is not weakened (disregarding secondary quasi-polarizing angle effects), hence the background remains strong but the multiply reflected beams again pass two silver films so that the reflected parallel system remains sharp, but weak, and is lost in the bright background.

The above explanation was proved to be correct by the following simple experiment. Light from a source (any kind) was passed through a single silver film on glass, and a Nicol prism interposed in the beam. The transmitted light was examined at varying angles of incidence. A surprising fact emerged. At normal incidence the transmitted light was unpolarized. As the incidence increased it became partially polarized, the polarization becoming more and more complete as the angle of incidence was increased, until when close to grazing the transmitted light was almost completely polarized, with the magnetic vector perpendicular to the plane of incidence. There appears to be no mention of this property of thin silver films in the literature. It is clear that the mechanism closely resembles that which makes a tourmaline plate into a polarizer. In both cases one of the vibrations is preferentially absorbed, but in the case of silver the preferential absorption depends upon the angle of incidence. The rapid decrease in intensity of the parallel vector is not to be attributed to the effective increase in the thickness of the film with growing incidence. This effect accounts for the steady slow weakening of the perpendicular component, but not for the rapid fall in the other component. A thin transparent silver film at variable incidence can be considered equivalent to tourmaline plates of variable thickness.

The Differential Phase Doubling of Fabry-Perot Fringes.

The previous discussion has been restricted to the Newton's rings localized fringes. The effect, of necessity, is also shown by the wedge fringes formed between two silvered glass flats. These show no special features, and can be employed equally well as the Newton's rings for the study of the phase change effects. In the discussion following, the influence of the differential phase change on fringes of *equal inclination* formed at infinity with silvered parallel plates (Fabry-Perot fringes) will be considered.

It is clear that if the proposed differential phase change explanation for the Newton's rings is correct, then a Fabry-Perot interferometer used at angles other than the normal should exhibit double fringes in an identical manner. It is a surprising fact that this has never been recorded, despite the many investigations made with the important Fabry-Perot interferometer. The effect was sought for and clearly demonstrated by the following procedure.

A Fabry-Perot interferometer with plate separation of average value (say 1 to 50 millimetres) is unsuitable for easily showing the effect of the phase change. To begin with, natural line width is resolved and hyperfine structure difficulties might be encountered. Then again, with such plate separations it is obvious that at high angles of incidence most of the multiply reflected beams are lost. For example, at even only 60° a beam has traversed *across* the face of the interferometer a distance $2\sqrt{3}$ times the plate separation after a double reflection. Taking a 1 millimetre gap interferometer of plate aperture 3 centimetres, then a beam incident

at the centre of the instrument at 60° is lost after only some four or five reflections. The loss of beams destroys the whole optical mechanism and coarse broad fringes result. Clearly then, on both the above counts, the thickness of the interferometer must be less than 0.1 millimetre if sharp fringes are to be seen at high angles of incidence. There is yet a further advantage in using so small a spacing. Very high order rings are seen at such angles and unless the gap is small the rings will be extremely close together. Since the linear ring dimensions are inversely as the square root of the plate separations, a thin gap leads to excellent fringe separations at high angles of incidence.

A Fabry-Perot interferometer was therefore set up with a small gap, the spacers being simply three pieces of writing paper, about $1/20$ mm. thick, adjusted to equality by screw pressure. On viewing the Fabry-Perot rings at infinity with a telescope, the expected doubling was observed to take place as the incidence was increased from the normal. The observed displacements and the changes in intensity were not measured, but clearly seemed to correspond identically with those found with Newton's rings, with the necessary difference that the weaker variable component moves *inwards* into the position of higher order of interference instead of outwards as with Newton's rings.

The influence of this doubling of the Fabry-Perot rings on high resolution spectroscopy will be discussed in Part III.

Applications of the New Interference Phenomena.

The phenomena described here have numerous obvious applications. These can broadly be divided off into uses with normal and with non-normal incidences. With normal incidence two features are evident, namely the increased sharpness and the increased intensity. The increased sharpness is such that the Newton's rings, as modified by silvering, can now be used with very high precision for metrological purposes. It is not too much to claim that very small displacements of the order of one-thousandth of a wave-length should be detectable with components of good optical quality. This far exceeds the accuracy possible with, say, Michelson's interferometer, the reason being of course the difference in the intensity distribution of the fringes in the two cases.

A matter of some considerable interest is that this procedure shows up the defects on the lens surface in a manner somewhat analogous to that employed by Fabry for flats, who used Fizeau fringes in transmission for this purpose. However, the Fizeau procedure is normally limited to *plane* surfaces which are nearly parallel. This procedure has always been recognized as giving the highest available precision. Now we have available a procedure for *curved* surfaces which offers the same precision exactly as formerly existed only for plane surfaces. The Newton's ring method of approach has the further advantage that at approximately normal incidence the degree of parallel collimation is by no means as critical as in the Fizeau arrangement, and indeed, with some tolerance permissible, an extended source can be employed.

The most important use of the fringes at non-normal incidence is clearly the fact that we have here available a new objective method for studying the optical properties of thin metallic films, in particular, the measurement of the differential phase change effects and the differential absorption effects. Earlier determinations of differential phase change involved the difficult evaluation of elliptic polarization. The present procedure shows up the phase change in an unmistakable objective manner. By enclosing thin films of liquid between the silvered surfaces it becomes clear that a considerable field has been opened up. Experiments on these lines are now being made.

It is further clear that with modern techniques for metallic deposition the dependence of the phase changes on the thickness of the metal film, the nature of the particular metal and the influence of the mode of deposition, can all now be studied.

Note on Earlier Related Observations.

On completion of the work described above a search was made through the literature to see whether earlier investigators (and their number is considerable) had (a) tried the effect of silvering on Newton's rings components, or (b) observed fringes of related sharpness. It was found that two observers have previously employed silvering, but the *essential condition*, that of employing a high reflecting coefficient coating, was *not* fulfilled.

Boulouch⁽⁴⁾, in a brief but striking note, described, 50 years ago, a repetition of Fizeau's classical observations on the variation of the visibility of the rings with order number, when using yellow sodium light as the source. He points out in his note (and this is the first mention in the whole of the literature on interferometry of so important a fact) that by silvering the surfaces at which interference occurs the fringes will be sharpened, and he showed that this must follow from the application of Airy's formula for multiple reflection in a film. He carried out the experiment, and despite the crude silvering technique available then, succeeded in resolving the sodium D lines. It is only correct to point out that this short note gives for the first time the fundamental optical principles of both the Fabry-Perot and the Lummer plate interferometers, and deserves wider recognition than it has received in the past from writers on optics.

Since the technique of thin film silver deposition was crude 50 years ago, Boulouch's original experiments did not reveal the powerful method he had in fact begun to apply. Thus, these early investigations remained neglected and undeveloped. They did, however, have an important indirect effect on the Fabry-Perot interferometer.

The second experimenter to employ silvering was Wood⁽⁵⁾. Wood states that by *half silvering* the glass components the rings can be obtained in transmission, and he photographed the fringes so formed, at normal incidence only.

It is to be noted that Wood explicitly refers to *half-silvering*. As a

result of this, so few are the multiple reflections that very little fringe sharpening takes place. Indeed, Wood does not remark on any appreciable ring sharpening, and of course little can be expected with half silvering. The essential different feature which characterizes the present investigations which have revealed the new phenomena lies in the employment of *high reflecting coefficient* silverings and *not* half silverings. The difference optically between the two procedures can be considered more or less approximately analogous to the difference in resolution attained with gratings with many and with few lines respectively.

Only two types of interference were found in the literature which can at all be considered to be related to those described here, namely, Lummer⁽⁶⁾ straight line fringes, and Boulouch's double fringes. The Lummer fringes are, strictly, an adaptation of Talbot's⁽⁷⁾ way of producing Herschel's fringes.

Lummer's straight line fringes are somewhat distantly related. These are formed when polarized light, incident at angles approaching total reflection, suffers multiple reflections at a thin air film enclosed between the hypotenuse faces of a pair of right-angled prisms in close juxtaposition. These straight line fringes rather remotely resemble the Newton's ring elliptical fringes in that both are very sharp because of effective very high reflecting coefficient, higher than available at the usual normal incidence. Roughly speaking, as ordinary Lummer plate fringes are to Fabry-Perot fringes, so are Lummer's straight line fringes to be compared with silvered Newton's rings at large angles of incidence.

Boulouch, in the brief note mentioned earlier, described double fringes which are in fact a special limited case of the general type discussed here. One face, the back face, of a thin film of glass was heavily silvered. The unsilvered face was then illuminated with monochromatic light. The back reflected fringes produced by interference at the silvered and the unsilvered faces were examined. At normal incidence, localized thin film fringes of moderate visibility were observed (the difference in reflecting power of the two faces is responsible for the poor visibility). On approaching *grazing* incidence, the back reflected fringes became much sharper because of the effective increase in the reflecting coefficient of the glass near to grazing incidence. At the same time the fringes were seen to double. This doubling is due to the differential phase change at reflection, but the effect is complicated by virtue of the fact that alternate reflections take place at glass and silver surfaces.

Since Boulouch of necessity could only detect the doubling *when close to grazing incidence*, no measurements could be made, nor could the dependence upon incidence be studied. It is clear that these double fringes are but a special case of the more general type described here in detail. Of particular interest is the very fact that Boulouch *could see two components* when viewing near to grazing incidence. In his arrangement the light reaching the observer does not pass *through* any silver film and no preferential absorption effect takes place. Thus the differently polarized beams emerge with comparable intensity.

APPENDIX.

Note on the Approximations Used in Part I.

In Part I. (Phil. Mag. xxxiv. p. 555 (1943)), which describes non-localized fringes formed with a Fabry-Perot interferometer, it is stated that the path difference between the first and the n th beams reaching a point on the screen to form the first ring is (p. 558)

$$2nT \cos \phi + 2n^2T(\lambda/L) \cos \phi.$$

It is considered advisable to show below how this is derived. It is necessary to note that care must be taken in making approximations, otherwise quantities comparable with the small path differences of the order of wave-lengths may easily be discarded, leading to gross errors. It is therefore *essential to approximate to the second order*. Using the notation of Part I., the path difference D can be written *exactly* as

$$\begin{aligned} D &= \sqrt{R^2 + L^2 + 4LT + 4T^2} - \sqrt{L^2 + R^2} \\ &= (L^2 + R^2)^{\frac{1}{2}} \left\{ \left(1 + \frac{4LT + 4T^2}{L^2 + R^2} \right)^{\frac{1}{2}} - 1 \right\} \\ &= (L^2 + R^2)^{\frac{1}{2}} \{ (1 + \alpha)^{\frac{1}{2}} - 1 \}, \end{aligned}$$

where

$$\frac{4LT + 4T^2}{L^2 + R^2} = \alpha.$$

Expanding the expression containing α to include the second order term gives

$$\begin{aligned} D &= (L^2 + R^2)^{\frac{1}{2}} (\alpha/2 - \alpha^2/8) \\ &= \frac{L}{\cos \phi} \left\{ \left(\frac{2LT + 2T^2}{L^2} \right) \cos^2 \phi - 2 \left(\frac{L^2T^2 + 2LT^3 + T^4}{L^4} \right) \cos^4 \phi \right\}. \end{aligned}$$

Powers of T/L greater than the second may justifiably be neglected, and if this is done we get

$$\begin{aligned} D &= \frac{L}{\cos \phi} \left\{ \frac{2T}{L} \cos^2 \phi + \frac{2T^2}{L^2} (\cos^2 \phi - \cos^4 \phi) \right\} \\ &= 2T \cos \phi + 2(T^2/L) \cos \phi \sin^2 \phi^*. \end{aligned}$$

The path difference D_n between the first and n th beam is obtained from this by replacing T by nT giving

$$D_n = 2nT \cos \phi + 2n^2(T^2/L) \cos \phi \sin^2 \phi.$$

For the first ring $D = 2T - \lambda$ so that closely enough $\sin^2 \phi = \lambda/T$, giving finally

$$D_n = 2nT \cos \phi + 2n^2T(\lambda/L) \cos \phi,$$

as given in Part I.

* In Part I. the index was omitted as a misprint from $\sin^2 \phi$. The error is *not* perpetuated in the succeeding derivations in Part I.

References.

- (1) Feussner, *Gehrcke's Handbuch der Physik. Optik.* vol. i. (1927).
- (2) MacLaurin, 'The Theory of Light' (1908).
- (3) Hamy, *Journal de Physique*, v. p. 789 (1906).
- (4) Boulouch, *Journal de Physique*, ii. p. 312 (1893).
- (5) Wood, 'Physical Optics' (1924).
- (6) Lummer, *Ann d. Physik.* xxii. p. 49 (1907).
- (7) Talbot, *Phil. Mag.* ix. p. 401 (1836).

XV. *Viscosities of Unsaturated Six-Membered Isocyclic Compounds.*

By J. NEWTON FRIEND and WILLIAM D. HARGREAVES*.

[Received December 10, 1943.]

THE objects of the present research have been (a) to study the effect on viscosity of progressive conversion of cyclohexane into benzene, and (b) to find the rheochor of the phenyl group.

Cyclic Hydrocarbons.

The viscosity of cyclohexane has already been studied ⁽¹⁾. Despite its lower density its viscosity is greater than that of benzene. It has been shown also that ring formation in cyclohexane is accompanied by a fall

TABLE I.

Densities, Viscosities and Rheochors at the Boiling Point.

	B. pt. °C.	D.	10 ³ η.	R.
Cyclohexane ⁽¹⁾ , C ₆ H ₁₂	80.7	0.7203	4.13	138.4
Cyclohexene, C ₆ H ₁₀	83.7	7483	3.35	126.7
1 : 3 Cyclohexadiene, C ₆ H ₈ ...	81.5	7727	3.11	118.6
Benzene ⁽²⁾ , C ₆ H ₆	80.2	8145	3.27	110.4

of 4.4 units in the rheochor. It appeared of interest, therefore, to study the effect of progressive unsaturation down to benzene. The results at the boiling point are summarized in Table I. ; the viscosities and densities over the whole range studied are shown graphically in fig. 1. These data are of special interest because, owing to the similarity of the boiling points, the temperature ranges are virtually identical.

Successive removal of each pair of H atoms causes a steady rise in density. On the other hand the viscosity falls, at first rapidly in hexene, then more slowly to a minimum in hexadiene, and rises again slightly in benzene. Irregularity in the behaviour of benzene is to be expected in view of the unique nature of its molecule.

* Communicated by the Authors.

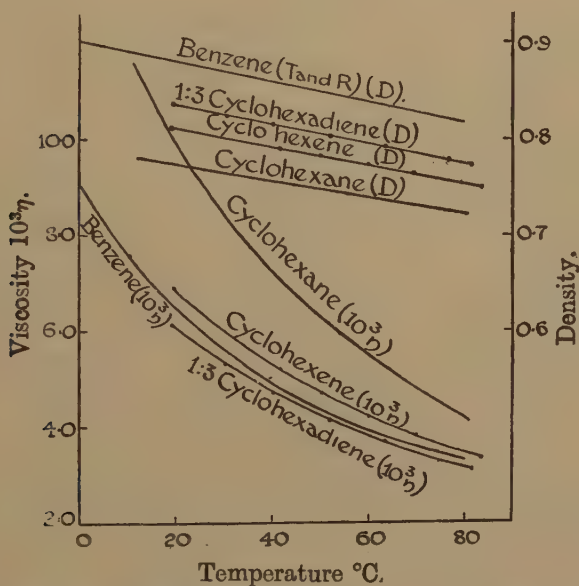
The rheochor of benzene is 0.6 unit greater than that estimated ⁽³⁾ as the sum of 6C+6H, the decrease due to ring formation being thus slightly more than counteracted by the effects of unsaturation.

The Phenyl Radicle.

As has already been shown ⁽¹⁾ ring formation in simple saturated derivatives of cyclohexane, such as cyclohexyl chloride, is accompanied on the average by a reduction of 5.6 units in the rheochor, that is 1.2 units greater than in the free hydrocarbon.

The effect of substitution in the benzene ring is even more pronounced, as one would expect in view of its unsaturation. The results are summarized in Tables II. and III. Densities and viscosities are shown graphically in figs. 2 and 3 over the temperatures studied.

Fig. 1.



D = Density curve.
 $10^3 \eta$ = Viscosity curve.

Diphenyl and diphenyl methane yield closely similar results. Introduction of one or more nitrogroups into benzene and toluene greatly increases the viscosity. On the other hand, the introduction of hydroxyl into benzene increases the viscosity even more than the nitro group, although Bramley's results show that orthonitrophenol is only slightly more viscous than phenol ⁽⁴⁾. It is hoped to deal more fully with ortho-effects in a later communication.

The phenyl ethers (fig. 3) show a progressive fall in density with increase in weight of the alkyl group. Their viscosities, however, rise with the molecular weight, but at their boiling points the viscosities of

the higher members approach each other. In these respects these ethers resemble the fatty ethers. Replacement of hydrogen in phenol by CH_3

TABLE II.
Boiling Point Data for Benzene Derivatives.

	D.	$10^3\eta$.	R.	$\text{R}(\text{C}_6\text{H}_5)$.
C_6H_6 ⁽²⁾	0.8145	3.27	110.4	104.9
$\text{C}_6\text{H}_5\text{CH}_3$ ⁽²⁾	0.7876	2.48	130.2	100.9
$\text{C}_6\text{H}_5\text{C}_2\text{H}_5$ ⁽²⁾	0.7619	2.33	153.6	100.5
$\text{C}_6\text{H}_4(\text{CH}_3)_2$ ⁽²⁾ } (Mean <i>o</i> , <i>m</i> , <i>p</i>)	153.8	100.7
$\text{C}_6\text{H}_5\text{CH}_2\text{NH}_2$ ⁽¹⁾	0.8366	2.83	144.9	100.5
Mean	100.7
$\text{C}_6\text{H}_5\text{NO}_2$	1.0105	3.24	140.3	101.4
$\text{C}_6\text{H}_4(\text{NO}_2)_2$	1.1503	4.21	173.8	101.5
$\text{C}_6\text{H}_5\text{NH}_2$ ⁽¹⁾	0.8774	3.60	123.9	103.3
$\text{C}_6\text{H}_4\text{CH}_3\text{NH}_2$ ⁽¹⁾ } (Mean <i>o</i> , <i>m</i> , <i>p</i>)	146.8	102.3
$\text{C}_6\text{H}_5\text{OH}$ ⁽⁴⁾	0.9260	4.65	122.5	102.5
$\text{C}_6\text{H}_5\cdot\text{C}_6\text{H}_5$	0.8385	2.56	205.1	2×102.6
$\text{C}_6\text{H}_5\text{CH}_2\text{C}_6\text{H}_5$	0.8180	2.59	229.6	2×102.9
$\text{C}_6\text{H}_5\text{COCH}_3$	0.8683	2.90	157.0	101.7
$\text{C}_6\text{H}_5\text{Cl}$ ⁽⁵⁾	0.9817	2.88	129.8	102.5
$\text{C}_6\text{H}_5\text{Br}$ ⁽⁶⁾	1.3070	3.565	139.9	104.1
$\text{C}_6\text{H}_5\text{I}$ ⁽⁶⁾	1.5662	3.85	153.0	105.4
Mean	102.7

TABLE III.
Boiling Point Data for Phenyl Ethers.

	B. pt. $^{\circ}\text{C}$.	D.	$10^3\eta$.	R.	$\text{R}(\text{C}_6\text{H}_5)$.
$\text{C}_6\text{H}_5\text{OH}$ ⁽⁴⁾	183	0.9260	4.65	122.5	102.5
$\text{C}_6\text{H}_5\text{OCH}_3$ ⁽¹⁾ ...	152	0.8635	2.735	141.0	101.7
C_2H_5	172	8130	2.48	167.0	103.9
$n\text{C}_3\text{H}_7$	184	7928	2.44	190.3	103.4
$n\text{C}_4\text{H}_9$	209	7668	2.27	215.0	104.3
$n\text{C}_5\text{H}_{11}$...	230	7447	2.19	240.6	106.1

in anisole is characteristic of the passage from an alcohol to an ether. Both density and viscosity are greatly reduced as the result, presumably, of elimination of the hydrogen bond.

In Table II. it is shown that the rheochor of the phenyl group attached to alkyl groups is fairly constant, yielding a mean value of 100.7. In benzylamine the NH_2 is attached to the alkyl group and apparently

has but little effect on the phenyl group. The mean ring effect, including unsaturation, is a decrease of 3.6 units, as opposed to an increase of 0.6 units in free benzene.

Fig. 2.

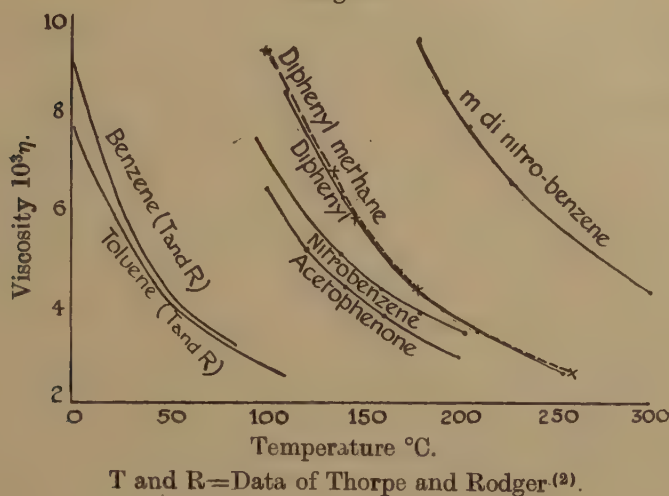
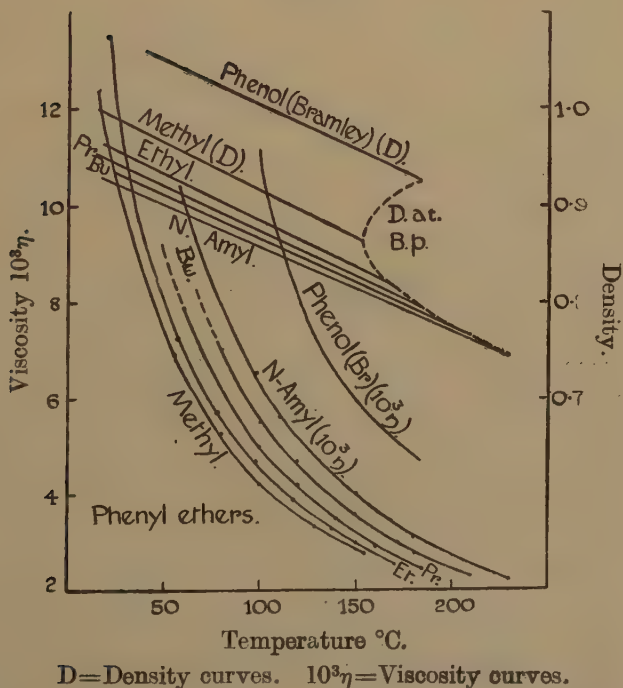


Fig. 3.



Replacement of hydrogen in benzene by an electronegative group, however, yields a higher value for the phenyl group, rising from 101.4 in nitrobenzene to 105.4 in iodobenzene. The mass of the substituent

group appears also to have an influence, as indicated by the ethers listed in Table III., in which the value rises from 101.7 in anisole to 106.1 in the amyl derivative.

Experimental.

Viscosities were determined with standard viscometers of the Ostwald type in the manner already indicated ⁽⁷⁾.

1 : 3 *Cyclohexadiene* was obtained from 1.2 dibromocyclohexane by treatment with quinoline ⁽⁸⁾. Two methods were employed in preparing the dibromocyclohexane, the first being that of Greengard ⁽⁹⁾. In the second method air was slowly aspirated through the calculated quantity of dried bromine and then through pure cyclohexane cooled in ice. In both cases the product boiled at 113° C. at 26 mm., but whereas the product by the former method darkened considerably on keeping (as stated by Greengard), the second method gave a colourless liquid which did not darken.

The *n*-propyl, butyl and amyl phenyl ethers were prepared by addition of 57 grams of phenol to 14 grams of sodium dissolved in 50 c.c. absolute alcohol. Slight excess of the alkyl bromide was added and the mixture refluxed until no longer alkaline. The alcohol was distilled off, the residue diluted with ether, washed and dried.

All the other substances used were obtained direct from British Drug Houses. Each was dried and distilled directly into the viscometer or density bottle, the middle fraction alone being used.

In all cases density values refer to water at 4° C., and any values quoted from literature have been re-calculated when necessary to render strictly comparable. The values of A and B refer to Sheppard's equation and to the linear part of the curve obtained on plotting $\log 10^3\eta$ against $1/T$. The boiling point values for η and D, obtained by extrapolation, are enclosed in parentheses.

Cyclohexene.—B. pt. 83.5° to 83.7° C. at 757.6 mm. $A = -0.8806$; $B = 501.4$, from 18°.

Temp.	18.7	41.6	50.0	60.2	69.75	79.6	83.7
D	0.8105	0.7892	0.7819	0.7716	0.7621	0.7525	(0.7483)
1000 η	6.893	5.180	4.720	4.231	3.821	3.471	(3.35)
$R = 126.7$.							

Evans ⁽¹⁰⁾ also determined the viscosity, but gave no density data; his value at 80°, viz. 3.37, is close to ours.

1 : 3 *Cyclohexadiene*.—B. pt. 81.0° to 81.5° C. at 745.7 mm. $A = -0.8790$; $B = 745.7$ mm., from 18°.

Temp.	18.85	30.4	40.3	52.3	63.6	76.9	81.5
D	0.8351	0.8242	0.8142	0.8022	0.7912	0.7776	(0.7727)
1000 η	6.114	5.287	4.704	4.128	3.676	3.236	(3.107)
$R = 118.6$.							

Our interpolated density at 25°, viz. 0.8290, is identical with that given by Perkin ⁽⁸⁾.

Nitrobenzene.—B. pt. 207.6° C. at 752.4 mm. $A = -0.6912$; $B = 577.5$, from 96°.

Temp.	96.4	139.5	160.9	180.0	199.5	207.6
D	1.1285	-1.0849	1.0625	1.0414	1.0197	(1.0105)
1000 η	7.447	5.084	4.351	3.837	3.394	(3.24)

$R = 140.3$

The viscosity found by Bingham ⁽¹¹⁾ at 100.5°, his highest temperature, viz. 7.045, lies reasonably close to our interpolated value 7.16.

m-Di-nitrobenzene.—B. pt. 300° C. with decomposition. $A = -0.6797$; $B = 747.1$, from 160°.

Temp. ...	100.0	129.5	160.5	178.0	193.0	205.0	227.0	300
D	1.3551	1.3248	1.2938	1.2761	1.2614	1.2489	1.2261	(1.1503)
1000 η	35.9	16.63	11.16	9.51	8.424	7.668	6.533	(4.21)

$R = 173.8$.

Diphenyl.—B. pt. 253° to 254° C. at 757.1 mm. $A = -0.9784$; $B = 730.4$, from 110°.

Temp.	110.5	139.5	154.5	170.0	182.0	193.3	210.0	254.0
D	0.9612	0.9376	0.9244	0.9111	0.9005	0.8905	0.8762	(0.8385)
1000 η	8.445	6.183	5.374	4.667	4.237	3.871	3.425	(2.56)

$R = 205.1$.

The viscosity found by Bingham ⁽¹²⁾ at his highest temperature, 100° C., was 9.709; our extrapolated value being 9.55.

Diphenyl methane.—B. pt. 260° to 260.5° C. at 757.8 mm. $A = -0.8539$; $B = 676.6$, from 130°.

Temp.	59.6	100.0	129.9	145.4	161.5
D	0.9757	0.9437	0.9200	0.9082	0.8950
1000 η	15.64	9.31	6.805	5.820	5.072

Temp.	178.3	191.3	211.0	260.5
D	0.8821	0.8711	0.8569	(0.8180)
1000 η	4.426	4.025	3.492	(2.59)

$R = 229.6$.

Acetophenone.—M. pt. 19.2°; B. pt. 201.3° C. at 756.5 mm. $A = -0.7689$; $B = 584.7$, from 100°.

Temp.	99.0	122.0	141.5	162.0	187.5	201.3
D	0.9596	0.9401	0.9224	0.9042	0.8808	(0.8683)
1000 η	6.349	5.165	4.393	3.788	3.184	(2.90)

$R = 157.0$.

Kurnakow ⁽¹³⁾ found the viscosity at 95° C. to be 6.53; our extrapolated value is 6.60.

Phenyl ethyl ether.—B. pt. 171.6° to 172.0° C. at 758.4 mm. $A = -0.9374$; $B = 594.5$, from 60°.

Temp.	18.0	57.8	77.6	97.1	117.0
D	0.9648	0.9277	0.9095	0.8908	0.8725
1000 η	13.56	7.246	5.738	4.673	3.899

Temp.	137.5	149.3	165.0	172.0
D	0.8509	0.8383	0.8208	(0.8130)
1000 η	3.261	2.970	2.619	(2.48)

$R = 167.0$.

Bramley's viscosities⁽⁴⁾ below 80° C. are somewhat lower than ours, but at 80°, his highest temperature, his figure, 5.58, is identical with our interpolated value.

Phenyl n-propyl ether.—B. pt. 184.2° C. at 753.7 mm. $A = -1.0770$; $B = 670.0$, from 140°.

Temp.	15.9	60.5	99.5	109.5	120.0	130.3
D	0.9521	0.9121	0.8766	0.8670	0.8575	0.8479
1000 η	17.27	7.872	5.054	4.592	4.179	3.833
Temp.	140.5	151.0	160.5	167.5	175.0	184.2
D	0.8379	0.8270	0.8173	0.8099	0.8023	(0.7928)
1000 η	3.476	3.178	2.940	2.780	2.620	(2.44)
$R = 190.3$.						

Phenyl n-butyl ether.—B. pt. 209.3° C. at 754.8 mm. $A = -1.0334$; $B = 671.2$, from 130°.

Temp.	22.95	80.7	101.3	110.0	120.0	129.5
D	0.9331	0.8832	0.8654	0.8580	0.8494	0.8416
1000 η	17.86	7.031	5.526	5.094	4.670	4.307
Temp.	140.4	149.0	162.0	172.5	182.8	195.0
D	0.8320	0.8236	0.8115	0.8017	0.7916	0.7802
1000 η	3.916	3.623	3.251	2.985	2.754	2.512
$R = 215.0$.						

Phenyl n-amyl ether.—B. pt. 229.5° to 230° C. at 756.4 mm. $A = -1.0411$; $B = 695$, from 130°.

Temp.	17.0	61.0	99.8	109.5	119.3	130.5	139.3
D	0.9304	0.8943	0.8617	0.8534	0.8450	0.8358	0.8287
1000 η	25.62	10.84	6.504	5.861	5.342	4.802	4.432
Temp.	150.3	165.5	181.2	195.0	210.0		229.5
D	0.8191	0.8051	0.7902	0.7773	0.7633		(0.7447)
1000 η	4.037	3.527	3.112	2.793	2.500		(2.19)
$R = 240.6$.							

References.

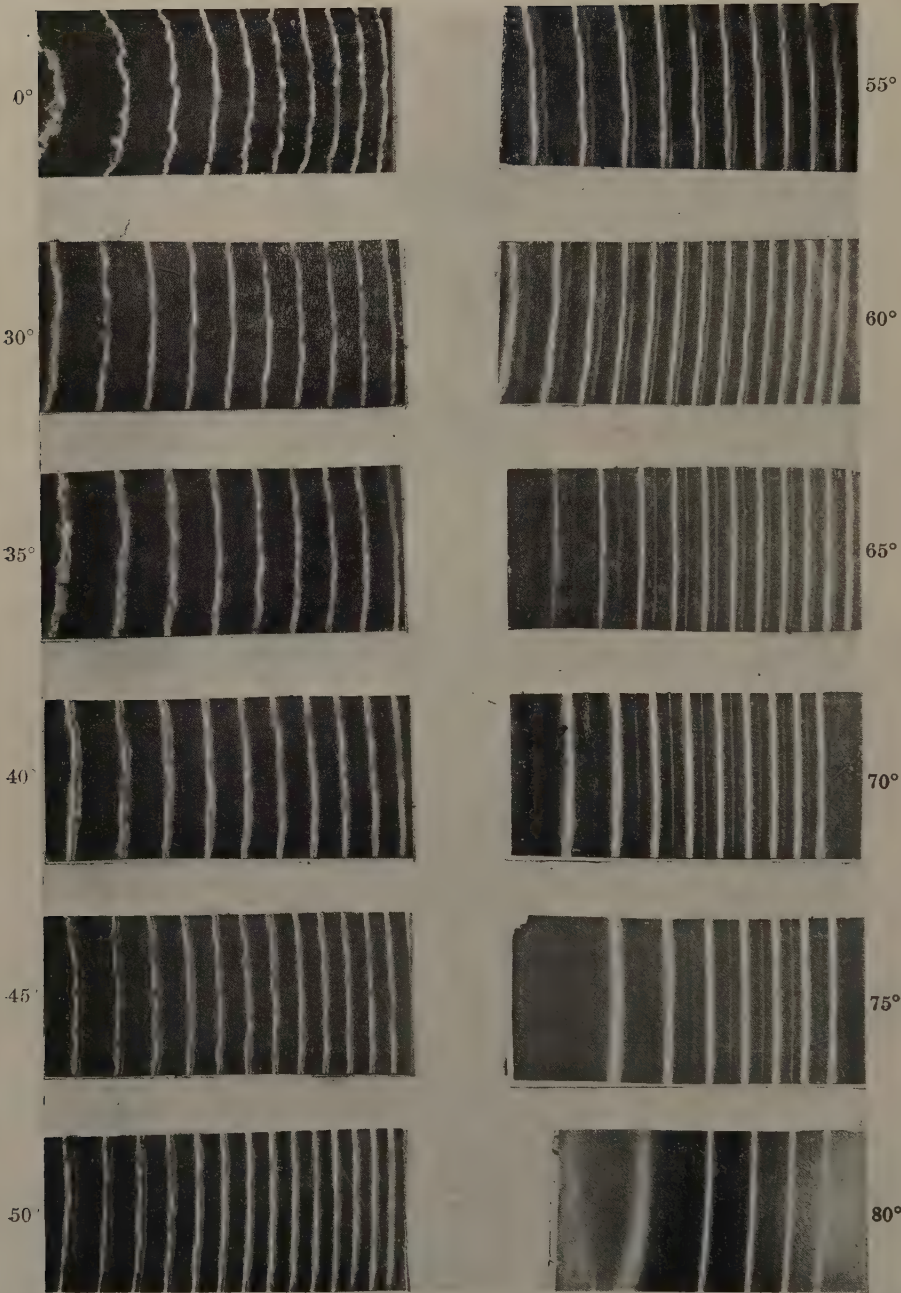
- (1) Friend and Hargreaves, *Phil. Mag.* xxxiv. p. 810 (1943).
- (2) Data of Thorpe and Rodger, *Phil. Trans.* clxxxv. p. 476 (1895).
- (3) Rheochors as follow: $C = 12.8$; $H = 5.5$.
- (4) Data of Bramley, *J. Chem. Soc.* cix. p. 10 (1916).
- (5) Data of Titani, *Bull. C. S. Japan*, ii. p. 95 (1927).
- (6) Data of Meyer and Mylius, *Z. physikal. Chem.* xcv. p. 349 (1920).
- (7) Friend and Hargreaves, *Phil. Mag.* xxxiv. p. 643 (1943).
- (8) Crossley (and Perkin), *J. Chem. Soc.* lxxxv. p. 1416 (1904).
- (9) Greengard, "Organic Syntheses," vol. xii. p. 26 (Wiley).
- (10) Evans, "II. Congr. mondiale petrole 2, Sec. 2. Phys. Chim. raffinage," p. 933 (1937).
- (11) Bingham, van Klooster and Kleinspehn, *J. Phys. Chem.* xxiv. p. 1 (1920).
- (12) Bingham, *J. Rheology*, i. p. 406 (1930).
- (13) Kurnakow, Krotkov and Oksman, *J. Russ. Chem. Soc.* xlvii. p. 583 (1915).

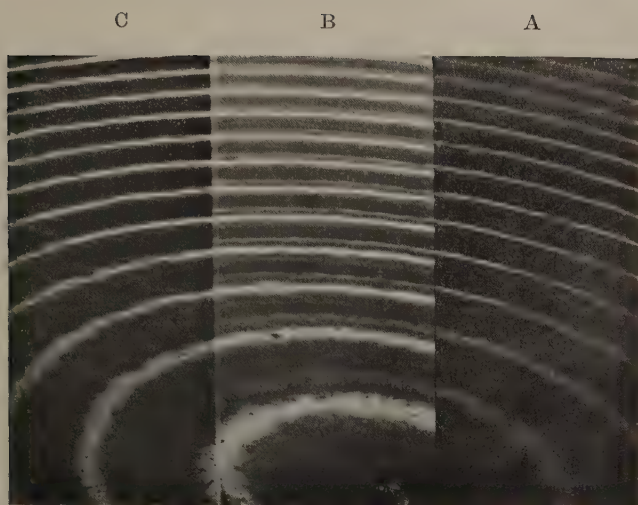
The Technical College,
Birmingham.

[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]



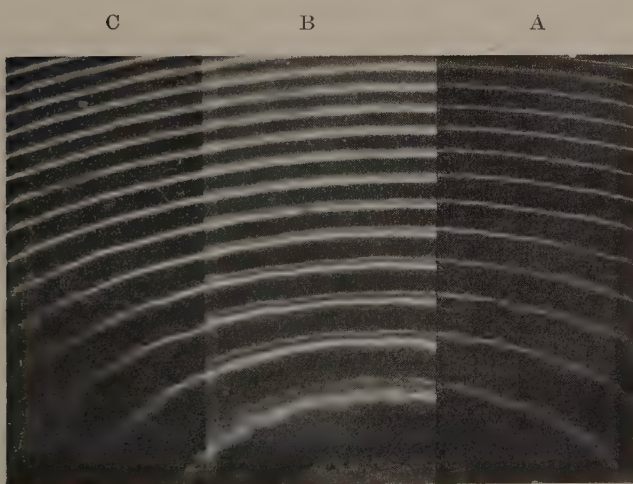
Multiple-beam transmission.
Newton's rings by normal incidence.





(a)

Nicol prism. 60° incidence.



(b)

Polaroid disc. 55° incidence.

XVI. *Transient Response in Frequency Modulation.*

By D. A. BELL, M.A., B.Sc., A.M.I.E.E.*

[Received May 7, 1943.]

SUMMARY.

Square-wave modulation is assumed to be applied both to an amplitude-modulated and a frequency-modulated system, in each of which the band-width is limited by a single resonant circuit, tuned to the carrier frequency, and the responses are computed. It is found that F.M. gives a better transient response than A.M. The response of a single tuned circuit to sinusoidal frequency modulation, of such high frequency of modulation that the side-bands fall on the skirt of the circuit resonance characteristic, is also examined and it is shown that severe harmonic distortion is introduced.

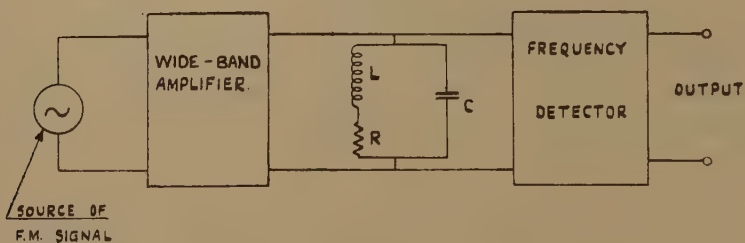
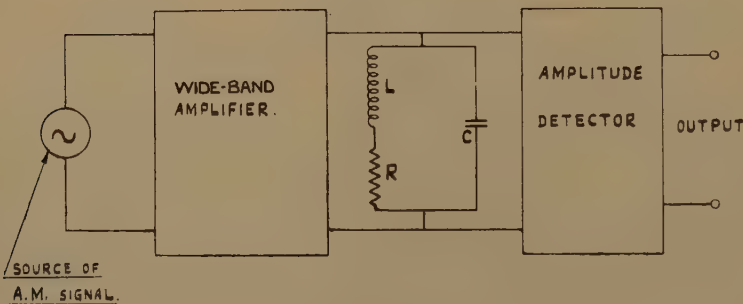
1. INTRODUCTION.

THE transient response of frequency-modulation systems has been discussed by Salinger ⁽¹⁾ in terms of a system using an "ideal filter," a hypothetical device of which the sole effect is to restrict the range of frequencies transmitted. The writer considered Salinger's analysis to be unsatisfactory, because it led to the result that the response to unit transient had already reached half-amplitude at $t=0$, and the first half of the transient response, for $t<0$ was evaluated by extrapolating a function which was initially stated to be valid only for $t>0$; it has, in fact, been shown in a recent paper ⁽²⁾ that the usual mathematical treatment of the "ideal filter" is not consistent with the properties of real filters. The treatment developed below for a single tuned circuit leads to a result different from Salinger's conclusion that the transient response of F.M. systems is similar to that of A.M. systems, and there is a physical reason for believing that the result obtained for a single tuned circuit is parallel to the results which would be obtained in more complex circuits. The response to a transient of any real circuit consists of two parts, one corresponding to the impressed E.M.F. and the other characteristic of the circuit, and faithful reproduction of transients requires that the first shall predominate over the second; but it is a feature of F.M. systems that they are more effective than A.M. systems in discriminating in favour of the stronger of two signals of different frequency, and they should, therefore, give a better transient response by discriminating more effectively in favour of the component corresponding to the impressed E.M.F. as against the free oscillation of the circuit.

* Communicated by the Author,

2. RESPONSE TO SQUARE-WAVE AMPLITUDE MODULATION.

The procedure in this paper is to compare the transmission of a square-wave signal through a frequency-modulated system with its transmission through an amplitude-modulated system, assuming for the sake of simplicity that in both systems the limitation on fidelity is imposed by a single parallel-resonant circuit. The two systems are then as shown in figs. 1*a* and 1*b*. For amplitude-modulation (fig. 1*b*) we have to find the response of the L, R, C circuit to a carrier wave (whose frequency may be assumed equal to the resonant frequency of the tuned circuit) modulated in amplitude by a square wave. Since the system is linear the response to square-wave modulation about any carrier level will be

Fig. 1*a*.Fig. 1*b*.

the same as the response to a similar square-wave modulation having its lower limit at zero amplitude, *i. e.*, to starting and stopping the carrier. This is merely the response to modulation by the Heaviside unit function, and it is known that at each transition the approach to the final magnitude takes the form

$$V = V_0 \left(1 - e^{-\frac{Rt}{2L}} \right) \dots \dots \dots (1)$$

In order to generalize this expression, note that $Q = \omega L/R$, where ω is the angular frequency alike of the applied signal and of resonance of the circuit; we then have

$$V/V_0 = 1 - e^{-\frac{\omega t}{2Q}}, \dots \dots \dots (1a)$$

which is independent of the separate values of Q and the scale of time. In fig. 2, V/V_0 is plotted against $\omega t/2Q$ from equation (1 a) as curve (a), and this curve represents the transient response which is typical of a single-circuit amplitude-modulated system.

3. RESPONSE TO SQUARE-WAVE FREQUENCY MODULATION.

3.1. Nature of the Signal.

Turning now to the frequency-modulated system, fig. 1 a, the signal must be assumed to be of constant amplitude, but to have its frequency abruptly changed according to the square-wave modulation. In order to use existing technique on the problem, this frequency-modulated signal may be represented in a way which has already been indicated elsewhere⁽³⁾: at each frequency transition it is supposed that a signal equal and opposite to the existing signal is applied, thus eliminating the previous frequency, and simultaneously another signal of equal amplitude but of the new frequency is applied. The net result is a change of frequency without change of amplitude, but it is expressed as the sum of two amplitude changes which can be dealt with by known technique. The resultant oscillation in the L, R, C circuit will consist of two components, one at the impressed frequency, and the other at the natural frequency of the circuit and having an exponential decay. Now, in a frequency-modulation system it is the frequency, not the amplitude, of the signal which conveys the modulation; although the "frequency detector" shown in fig. 1 a may, in practice, consist of an amplitude-limiter followed by a frequency-selective network (discriminator), and then an amplitude detector, the only requirement of the corresponding analytical process is that it shall determine the instantaneous frequency of the resultant oscillation in the circuit, regardless of its amplitude.

3.2. Determination of Instantaneous Frequency of Complex Wave.

If an oscillation is of the form $\sin(\omega t + \phi)$, where ϕ may be a function of t , the instantaneous frequency is defined as

$$\bar{\omega} = \frac{d}{dt} \{\omega t + \psi\} = \omega + \frac{d\psi}{dt} \quad \dots \quad (2)$$

If we have two oscillations, $R \sin \omega_1 t$ and $S \sin \omega_2 t$, it can be shown⁽⁴⁾ that the instantaneous frequency is

$$\omega = \frac{\omega_1 R^2 + \omega_2 S^2 + (\omega_1 + \omega_2) RS \cos(\omega_1 - \omega_2)t}{R^2 + S^2 + 2RS \cos(\omega_1 - \omega_2)t} \quad \dots \quad (3)$$

Equation (3) gives the instantaneous frequency of the resultant of any two sinusoidal oscillations; in the application later in this paper, one of the two will be the forced oscillation and the other the free oscillation set up by the transient. Strictly speaking, further terms should be added to (3) when either R or S is a function of t , as it will be when one

of them is a free oscillation decaying exponentially; but in practice the decay constant is so small compared with ω that these additional terms may be neglected.

Since $\bar{\omega}/\omega_1$ differs only slightly from unity, it is advantageous to express it in the form $1+\delta$; therefore equation (3) is rewritten as

$$\frac{\bar{\omega}}{\omega_1} = 1 + \left(\frac{\omega_2}{\omega_1} - 1 \right) \left\{ \frac{S^2 + RS \cos (\omega_1 - \omega_2)t}{R^2 + S^2 + 2RS \cos (\omega_1 - \omega_2)t} \right\}. \quad (3a)$$

3.3. *Output Frequency during an Amplitude Transient.*

As explained in section 1, the F.M. transient (abrupt change from one frequency to another) is to be expressed as the sum of two amplitude transients, so it is first necessary to determine the response of the L, R, C circuit to the sudden injection of a frequency other than the resonant frequency. A convenient way of determining this response is the operational method, and in particular the use of a contour integral to interpret the operational expression obtained. (This method was suggested to the author by Mr. A. H. Wynn.)

It will be assumed that the amplifier of fig. 1a delivers a constant current to the tuned circuit (*e.g.*, from a pentode amplifier), and that the frequency detector is a device of infinite input impedance detecting the frequency of the voltage developed across the tuned circuit. (The finite input impedance of an actual frequency detector can, of course, be taken into account by modification of the values of C and R in the tuned circuit.)

The operational form of the impedance of the parallel resonant circuit is

$$Z(p) = \frac{1}{C} \left\{ \frac{p + 2b}{(p + b)^2 + \omega^2} \right\}, \quad \dots \quad (4)$$

where $b = R/2L$ and $\omega^2 = 1/LC - R^2/4L^2$. We wish to determine the response of the circuit to a suddenly applied current $i \sin \omega t$, the operational form of which current is

$$i(p) = i \cdot \frac{xp}{x^2 + p^2}.$$

The operational form of the resulting voltage is therefore

$$V(p) = \frac{i}{C} \left\{ \frac{p + 2b}{(p + b)^2 + \omega^2} \right\} \left\{ \frac{xp}{x^2 + p^2} \right\}. \quad \dots \quad (5)$$

This can now be interpreted into a direct expression of V as a function of time by the use of a contour integral (see, for example, Bush, 'Operational Circuit Analysis,' chapter ix.), and it is finally possible to express the whole response of the circuit in the form

$$V(t) = \frac{i}{\omega C} \left\{ A \cos \omega t - B e^{-\frac{\omega t}{2Q}} \cos \omega t + C \sin \omega t - D e^{-\frac{\omega t}{2Q}} \sin \omega t \right\}, \quad (6)$$

where

$$\left. \begin{aligned} A &= \frac{\omega}{x} \left[\frac{(x^2/\omega^2)(1-x^2/\omega^2+1/4Q^2)-x^2/\omega^2Q^2}{(1-x^2/\omega^2+1/4Q^2)^2+x^2/\omega^2Q^2} \right], \\ B &= \frac{\omega}{x} \left[\frac{(x^2/\omega^2)(1-x^2/\omega^2+1/4Q^2)-x^2/\omega^2Q^2}{(1-x^2/\omega^2-1/4Q^2)^2+x^2/\omega^2Q^2} \right], \\ C &= \frac{\omega}{x} \left[\frac{x^3/\omega^3Q+(x/\omega Q)(1-x^2/\omega^2-1/4Q^2)}{(1-x^2/\omega^2+1/4Q^2)^2+x^2/\omega^2Q^2} \right], \\ D &= \frac{\omega}{x} \left[\frac{x^2/\omega^2Q+(x^2/2\omega^2Q)(1-x^2/\omega^2-1/4Q^2)}{(1-x^2/\omega^2-1/4Q^2)^2+x^2/\omega^2Q^2} \right]. \end{aligned} \right\} \dots \dots (7)$$

It will be noted that (6) and (7) are in a generalized form: $i/\omega c$ is the product of the applied current and the reactance of the tuning capacity, x/ω indicates the extent of the departure of the applied frequency from the natural frequency of the circuit, b/ω is equal to $1/2Q$, and in (6) the exponential decay has been expressed in terms of $\omega t/2Q$, as was done in equation (1a).

If $x=\omega$, then $A=B$, and $C=D$ except for a term of order $1/Q^3$, which is negligible. In the more general case, $x \neq \omega$, it is still true that $A=B$, since the numerator of each reduces to $(x^2/\omega^2)(1-x^2/\omega^2-3/4Q^2)$; but the terms C and D , which are due to the resistance in the tuned circuit, do not match exactly. We may put these in the form

$$C = \frac{\omega}{x} \left[\frac{(x/\omega Q)(1+1/4Q^2)}{(1-x^2/\omega^2+1/4Q^2)^2+x^2/\omega^2Q^2} \right] \div \frac{(1+1/4Q^2)/Q}{1/16Q^4+1/Q^2}, \quad (8)$$

$$D = \frac{\omega}{x} \left[\frac{(x^2/2\omega^2Q)(2-x^2/\omega^2-1/4Q^2)}{(1-x^2/\omega^2-1/4Q^2)^2+x^2/\omega^2Q^2} \right] \div \frac{(1-1/8Q^2)/Q}{1/16Q^4+1/Q^2}, \quad (9)$$

using the approximation $x/\omega \doteq 1$. If the Q of the circuit is large, $1/4Q^2$ is negligible compared with unity, so $C=D$. With this assumption that C and D are equal, equation (6) can be simplified to

$$V(t) = K \left\{ \cos(xt - \phi) + e^{-\frac{\omega t}{2Q}} \cos(\omega t - \phi) \right\} \dots \dots (10)$$

It should be noted that A and B change sign when x changes from above to below the natural frequency ω of the circuit, but to the degree of approximation adopted above, for Q large, C and D remain of the same sign.

3.4. Output Frequency during a Frequency Transient.

Consider first the changes of frequency which accompany the amplitude transient when the input of frequency $x/2\pi$ is suddenly applied to a circuit of reasonably high Q . There will be two components present, as shown in equation (10), and the instantaneous frequency is therefore to be found with the aid of equation (3), which, with the values from (10), becomes

$$\omega = \frac{x + \omega e^{-\frac{\omega t}{Q}} + (x + \omega) e^{-\frac{\omega t}{2Q}} \cos(x - \omega)t}{1 + e^{-\frac{\omega t}{Q}} + 2e^{-\frac{\omega t}{2Q}} \cos(x - \omega)t} \dots \dots (11)$$

The modulation is conveyed by the change of frequency from the normal value (it is assumed that the natural frequency of the circuit coincides with the carrier frequency of the F.M. signal), so that the modulation would ideally be $x - \omega$, but is actually $\bar{\omega} - \omega$; expressed in fractional form, the relative response of the circuit to the frequency modulation is $(\bar{\omega} - \omega)/(x - \omega)$, a quantity which will be denoted by M/M_0 . In order to keep to generalized parameters, it will be written as

$$\frac{M}{M_0} = \frac{\bar{\omega}/x - \omega/x}{1 - \omega/x}, \dots \dots \dots (12)$$

and equation (11) can be re-arranged as

$$\frac{\bar{\omega}}{x} = \frac{1 + \omega/x e^{-\frac{\omega t}{Q}} + (\omega/x) e^{-\frac{\omega t}{2Q}} \cos(x - \omega)t}{1 + e^{-\frac{\omega t}{Q}} + 2e^{-\frac{\omega t}{2Q}} \cos(x - \omega)t} \dots \dots \dots (11a)$$

In wide-band frequency modulation the difference frequency $x - \omega$ may be so great that it is outside the range of the audio-frequency amplifier; in that case the terms in $\cos(x - \omega)t$ may be neglected (since their average value is zero) and equations (11a) and (12) reduce to the very simple expression

$$M/M_0 = 1 / \left(1 + e^{-\frac{\omega t}{Q}} \right) \dots \dots \dots (13)$$

This has the value of 0.5 at $t=0$, in contrast to the A.M. response (equation (1a)) which has the value zero at $t=0$. This result may seem surprising, but there is only one doubt, namely, that we have only examined the *frequency* of the resultant oscillation, and a real receiver will not respond unless the resultant has also a finite amplitude, preferably an amplitude sufficient to operate the limiter so that the receiver is able to respond to frequency modulation while rejecting amplitude modulation. In the notation of equation (3) the resultant of two oscillations has amplitude

$$\{R^2 + S^2 + 2RS \cos(\omega_1 - \omega_2)t\}^{\frac{1}{2}}.$$

When $R=S$, this oscillates between 0 and $2R$, but the discussion has so far been limited to the case where variations at the frequency $\omega_2 - \omega_3$ are averaged out, so that the effective amplitude is always finite. Equation (13) is plotted in fig. 2 as curve (b). This is not a true frequency-modulation transient, but the transient modulation of frequency which accompanies an amplitude transient. We now proceed to consider a symmetrical frequency-modulation transient, consisting of the sudden change of the applied frequency from x_2 to x_1 , where

$$1 - x_2^2/\omega^2 = x_1^2/\omega^2 - 1$$

by the method outlined in section 1. Using the nomenclature of equations (6) and (7), the steady state before the transient is represented by

$$(i/\omega c)\{A_2 \cos x_2 t + C_2 \sin x_2 t\} \dots \dots \dots (14a)$$

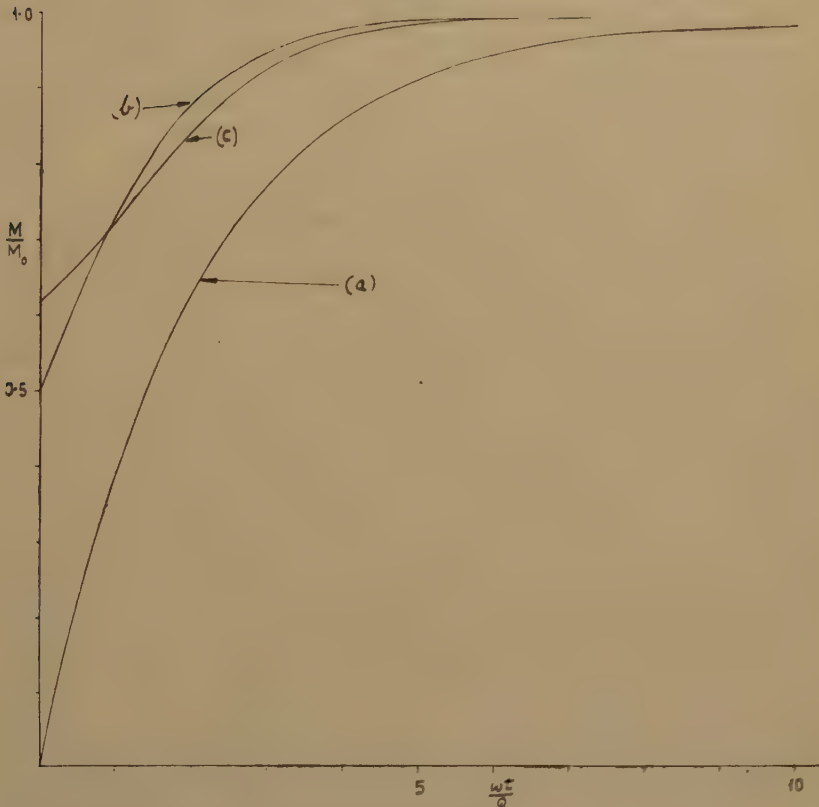
The sudden application of an oscillation equal and opposite to this will cause a response

$$(i/\omega C) \left\{ -A_2 \cos x_2 t + B_2 e^{-\frac{\omega t}{2Q}} \cos \omega t - C_2 \sin x t + D_2 e^{-\frac{\omega t}{2Q}} \sin \omega t \right\}. \quad (14b)$$

The sudden application of an oscillation at the new frequency will yield

$$(i/\omega C) \left\{ A_1 \cos x_1 t - B_1 e^{-\frac{\omega t}{2Q}} \cos \omega t + C_1 \sin x_1 t - D_1 e^{-\frac{\omega t}{2Q}} \sin \omega t \right\}. \quad (14c)$$

Fig. 2.



- (a) Amplitude response to amplitude transient.
- (b) Variation in frequency in the response to an amplitude transient.
- (c) F.M. response to transient modulation, with frequency sweep equal to twice circuit band-width.

Adding these three components (14 a), (14 b) and (14 c), the resultant response to the change is

$$(i/\omega C) \left\{ A_1 \cos x_1 t + (B_2 - B_1) e^{-\frac{\omega t}{2Q}} \cos \omega t + C_1 \sin x t + (D_2 - D_1) e^{-\frac{\omega t}{2Q}} \sin \omega t \right\}. \quad (15)$$

Taking into account the rules of sign at the end of section 3, and the assumed symmetry of the transmission which makes the magnitudes of A_1 , A_2 , etc., equal,

$$V = (i/\omega C) \left\{ A_1 \cos x_1 t + C_1 \sin x_1 t - 2B e^{-\frac{\omega t}{2Q}} \cos \omega t \right\}. \quad (16)$$

Evaluating the frequency $\bar{\omega}$ of the resultant of the components A_1 , $-2B_1$, C_1

$$\frac{\bar{\omega}}{x_1} = \frac{A_1^2 + C_1^2 + 4B_1^2(\omega/x_1) e^{-\frac{\omega t}{Q}} - 2B_1\{A_1^2 + C_1^2\}^{\frac{1}{2}}(1 + \omega/x_1) e^{-\frac{\omega t}{2Q}} \cos(\omega - x_1)t}{A_1^2 + C_1^2 + 4B_1^2 e^{-\frac{\omega t}{Q}} - 4B_1\{A_1^2 + C_1^2\}^{\frac{1}{2}} e^{-\frac{\omega t}{2Q}} \cos(\omega - x_1)t} \quad (17)$$

With the same assumption as before, that the difference frequency $\bar{\omega} - x_1$ is outside the audio-frequency band of the receiver, equation (17) simplifies to

$$\frac{\bar{\omega}}{x_1} = \frac{A_1^2 + C_1^2 + 4B_1^2(\omega/x_1) e^{-\frac{\omega t}{Q}}}{A_1^2 + C_1^2 + 4B_1^2 e^{-\frac{\omega t}{Q}}}. \quad (18)$$

Using equation (3a), equations (17) and (18) can be rewritten

$$\frac{\omega}{x_1} = 1 + \left(\frac{\omega}{x_1} - 1 \right) \left\{ \frac{4B_1^2 e^{-\frac{\omega t}{Q}} - 2B_1(A_1^2 + C_1^2)^{\frac{1}{2}} e^{-\frac{\omega t}{2Q}} \cos(\omega - x_1)t}{A_1^2 + C_1^2 + 4B_1^2 e^{-\frac{\omega t}{Q}} - 4B_1(A_1^2 + C_1^2)^{\frac{1}{2}} e^{-\frac{\omega t}{2Q}} \cos(\omega - x_1)t} \right\}. \quad (17a)$$

$$\text{and} \quad \omega/x_1 = 1 + (\omega/x_1 - 1) \left(4B_1^2 e^{-\frac{\omega t}{Q}} \right) / \left(A_1^2 + C_1^2 + 4B_1^2 e^{-\frac{\omega t}{Q}} \right). \quad (18a)$$

Ideally the modulation would be $x_2 - x_1$, but actually it is $x_2 - \bar{\omega}$, so that (remembering that $x_2 - \omega = \omega - x_1 = \frac{1}{2}(x_2 - x_1)$)

$$\frac{M}{M_0} = \frac{x_2 - \bar{\omega}}{x_2 - x_1} = \frac{1}{2} \left\{ 1 + \frac{\omega/x_1 - \bar{\omega}/x}{\omega/x_1 - 1} \right\}. \quad (19)$$

To evaluate (19) the constants A, C, and B which determine $\bar{\omega}/x$ must be found, or, since they all have a common denominator (to a close approximation) and a common factor ω/x , it is sufficient to work in terms of A' , B' and C' , the numerators of the fractions in equation (7)

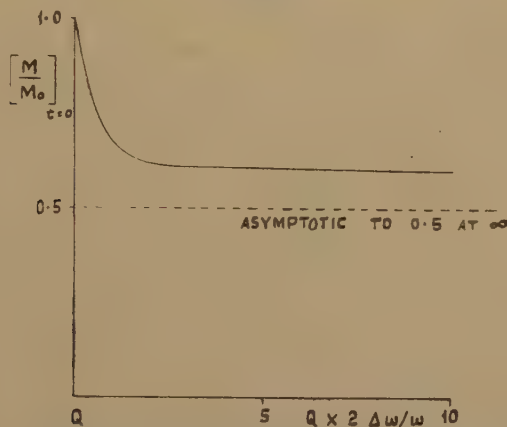
$$\left. \begin{aligned} A' &= B' = (x^2/\omega^2)(1 - x^2/\omega^2 - 3/4Q^2), \\ C' &= (x/\omega Q)(1 + 1/4Q^2) \div x/\omega Q. \end{aligned} \right\} \quad (20)$$

It has already been remarked that $1 - x^2/\omega^2 \div 2\Delta f/f$ and $1/Q$ is a measure of the band-width, so M/M_0 can be expressed as a family of curves each of which corresponds to a specific value of $Q \times 2\Delta f/f$.

In the limiting case of a very small shift of applied frequency, $\omega/x_1 \rightarrow 1$, the initial output frequency of the circuit approaches equality

with the applied frequency, so that $M/M_0 \rightarrow 1$. At the other extreme, very wide frequency deviation, the value of M/M_0 is asymptotic to 0.5. This latter result is obtained because change of applied frequency from $\omega + \Delta\omega$ to $\omega - \Delta\omega$ (where $\Delta\omega$ is large compared with the band-width of the circuit) results in setting up a strong transient oscillation at the natural frequency of the circuit; the output frequency then changes from $\omega + \Delta\omega$ to ω , instead of from $\omega + \Delta\omega$ to $\omega - \Delta\omega$, resulting in an initial response equal to only 50 per cent. of the applied modulation. In ordinary amplitude-modulation working, a commonly adopted convention is that the frequency-band covered by the circuit, expressed as a fraction of the mean frequency, is equal to the reciprocal of the Q of the circuit. The curve for F.M. with *twice* this frequency sweep has been plotted against a time scale of $\omega t/Q$ in the curve (C) in fig. 2. A whole family of such curves could be plotted, for various ratios of frequency

Fig. 3.



Initial frequency change when a square-wave frequency-modulated signal is applied to a resonant circuit.

sweep to circuit band-width; but for economy of labour the initial values only (corresponding to $t=0$ and neglecting terms in $\cos(x - \omega)t$) have been evaluated, and are plotted in fig. 3 against a scale of $Q \times 2\Delta\omega/\omega$.

4. PHYSICAL INTERPRETATION OF RESULTS.

The curves of fig. 2 show only a single transition, but by drawing a succession of these curves in alternate polarity, we can construct the response to a square wave, which is illustrated in fig. 4; curve (a) shows the ideal square wave, which represents the applied modulation; curve (b) shows the response of the A.M. system, and curve (c) the response of the F.M. system, having the same circuit band-width and a frequency deviation which is *large compared with the band-width of the circuit*, so that M/M_0 has its minimum value of 0.5, the value below which it cannot fall for the reason explained at the end of section 4. The F.M. and

amplitude cases are considered comparable because the relation of the circuit band-width to the period of the square-wave modulation is the same for each.

It is apparent from fig. 4 that the performance of the F.M. system is considerably better than that of the A.M. system, and the question arises as to how this is physically possible when the intelligence is passed through a circuit of the same band-width in either case. The answer is probably that in the F.M. case a great deal of possible signal energy is lost in the limiter*, and by accepting an output of less amplitude than would otherwise be obtained, the response to square-wave modulation can be improved. It is important to note, however, that the F.M. curves plotted are curves of *mean* frequency, superimposed on which there is, in fact, an oscillation at the difference frequency $\Delta\omega$. As long as the frequency modulation is sufficiently wide-band, $\Delta\omega$ may be

Fig. 4.



(a) Ideal square-wave modulation.



(b) Response of A.M. system.



(c) Response of F.M. system.

outside the band of modulation frequencies, and can therefore be eliminated subsequently. But it is necessary to ensure that when $\Delta\omega$ is small enough to be within the modulation-frequency range (as, for example, at low depths of modulation even in a wide-band F.M. system) the oscillation is of small amplitude; this means that the initial value of M/M_0 should not be much less than unity for these frequency sweeps; in other words, the circuit must have a band-width of the order of the range of modulation frequencies. A further point is that the initial change of frequency is not strictly instantaneous, but occupies a time of the order of a quarter period of $\Delta\omega$; but since taking the mean value

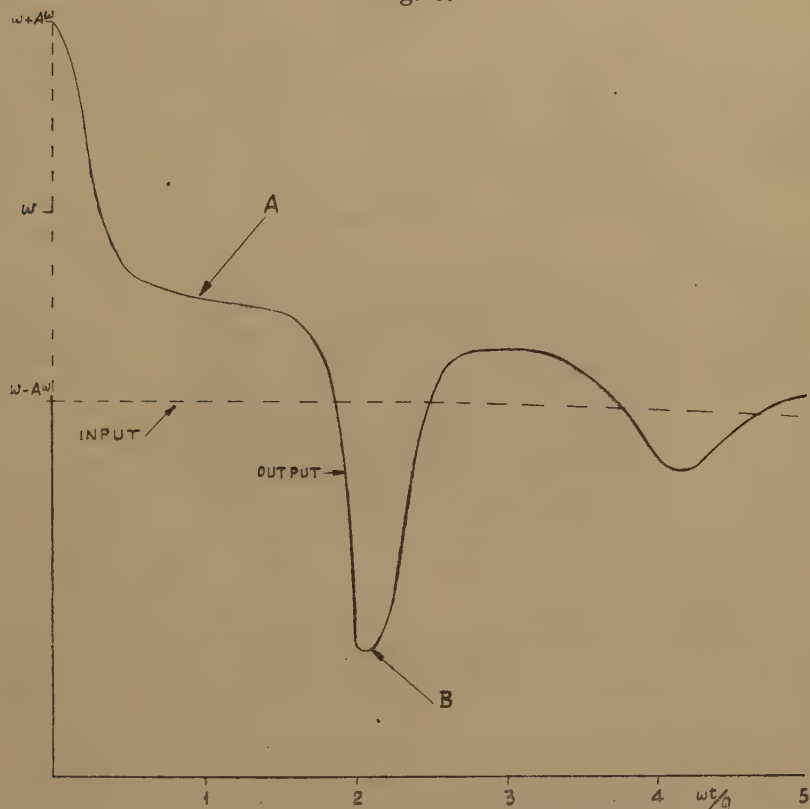
* Although the limiter characteristic is not explicitly shown in the analysis, it was pointed out in section 3.1 that the analytical procedure of examining the instantaneous frequency and ignoring the amplitude implies the assumption of a perfect limiter and frequency discriminator.

is equivalent to assuming that the signal will subsequently be passed through a circuit having a time-constant longer than the period of $\Delta\omega$, this is not important.

5. A PRACTICAL EXAMPLE.

For the sake of completeness, however, one case has been worked out in detail, and the response indicated by equation (17 *a*) is plotted in fig. 5. The constants assumed were circuit $Q=100$ and frequency sweep $=\pm 3$ per cent., so that the frequency sweep is several times greater

Fig. 5.



Detailed response of circuit to F.M. transient,
including component of frequency $\Delta\omega$.

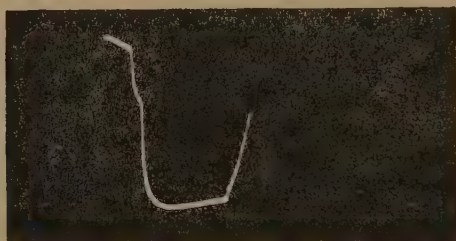
than the conventional band-width of the circuit. An attempt was made to examine this case experimentally, owing to the great practical importance of the conclusion already arrived at theoretically, that a circuit of band-width less than the frequency sweep of the F.M. signal causes little distortion other than a ripple at supersonic frequency. In order to have sufficient gain to operate a limiter, the I.F. amplifier (2 Mc./s) and detector of an existing F.M. receiver were used, and a sharply-tuned circuit placed between the signal generator and the input of the I.F.

amplifier. This tuned circuit had a band-width (at 3 db. loss) of 16 kc./s, and the signal-generator was frequency-modulated over a sweep of ± 60 kc./s by a square-wave generator adjusted to perform a complete cycle in about 1 millise. The receiver itself showed limitations under this test, but the only perceptible effect of inserting the sharply-tuned circuit in front of it was to accentuate an existing irregularity so as to produce the sharp kink in the side of the wave which can be seen in the oscillogram reproduced in fig. 6. This probably corresponds to the shoulder A of the calculated curve, the sharp peak B of the calculated curve being suppressed in the receiver owing to its short duration.

6. DISCRIMINATOR REQUIREMENTS.

So far the frequency detector has only been considered symbolically as a device capable of responding perfectly to changes of instantaneous frequency, while disregarding changes of amplitude. But, in practice, the frequency detector will usually consist primarily of a discriminator

Fig. 6.



Oscillogram of receiver output.

built up of a system of resonant circuits, whose output amplitude varies with frequency; consequently, the transient *amplitude* response of the discriminator circuits must be adequate.

7. COMPARISON WITH SIDE-BAND METHOD OF ANALYSIS FOR SINUSOIDAL FREQUENCY MODULATION.

The side-band spectrum for square-wave frequency modulation has been given by van der Pol ⁽⁵⁾, but the determination of the spectrum is the least part of the problem of finding the frequency-modulation response of a tuned circuit. It is necessary to find the amplitude and phase of each one of the side-bands after passing through the tuned circuit, and then combine all the side-bands into a single resultant, so that the resultant frequency can be found. (It will be realized that the limiter in an F.M. receiver cannot affect the amplitudes of the side-bands individually, but only the amplitude of the resultant; hence, although the limiter flattens the top of the amplitude-frequency characteristic of the I.F. amplifier as examined with a signal generator, it does not influence

the relative amplitudes of the side-bands in the Fourier representation of a frequency-modulated wave.)

If a carrier of angular frequency ω , with side-bands $\omega \pm \Delta\omega$ is applied to a tuned circuit, and the response of the circuit to the carrier be taken as unity, the response to the side-bands will have amplitude

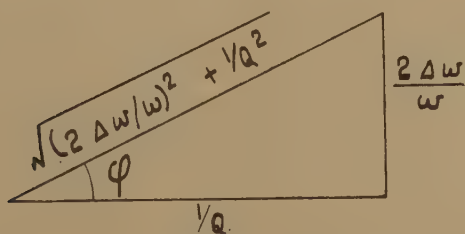
$$\frac{1/Q}{\{(2\Delta\omega/\omega)^2 + 1/Q^2\}^{\frac{1}{2}}} \quad \text{and} \quad \text{phase shift } \pm \tan^{-1}\left(Q \cdot \frac{2\Delta\omega}{\omega}\right). \quad (21)$$

Thus, if ϕ represents the phase shift and y the attenuation, the diagram of fig. 7 shows that $y = \cos \phi$. (This relationship enables trigonometrical tables to replace arithmetic in finding the attenuation of a series of uniformly spaced side-bands whose phase shift is also to be determined.) To keep the number of terms reasonably small, consider a sinusoidally frequency-modulated wave with a modulation-index of 2, so that the carrier and side-band amplitudes are given by the following factors:—

$$\begin{aligned} J_0(2) &= 0.2239, & J_1(2) &= 0.5762, & J_2(2) &= 0.3528, \\ J_3(2) &= 0.1289, & J_4(2) &= 0.0340, & J_5(2) &= 0.0070, \text{ etc.} \end{aligned}$$

A reasonable approximation can be obtained by stopping at $J_3(2)$.

Fig. 7.



If the band-width of the tuned circuit is such that for the first side-band $Q(2\Delta\omega/\omega) = 1$, the phase-shift for the various side-bands will be

$$\phi_1 = \tan^{-1} 1 = 45^\circ, \quad \phi_2 = \tan^{-1} 2 = 63.4^\circ, \quad \phi_3 = \tan^{-1} 3 = 71.6^\circ.$$

The corresponding attenuations are

$$y_1 = 0.7071, \quad y_2 = 0.4473, \quad y_3 = 0.3162.$$

If the original frequency-modulated wave was represented by

$$\begin{aligned} \sin(\omega t + 2 \sin ft) &= 0.2239 \sin \omega t + 0.5762 [\sin(\omega + f)t - \sin(\omega - f)t] \\ &+ 0.3528 [\sin(\omega + 2f)t + \sin(\omega - 2f)t] + 0.1289 [\sin(\omega + 3f)t \\ &- \sin(\omega - 3f)t], \quad \dots \quad (22) \end{aligned}$$

the resultant wave after passing through the tuned circuit will be

$$\begin{aligned} E' &= 0.2239 \sin \omega t \\ &+ 0.4074 \{\sin[(\omega + f)t - \phi_1] - \sin[(\omega - f)t + \phi_1]\} \\ &+ 0.1987 \{\sin[(\omega + 2f)t - \phi_2] + \sin[(\omega - 2f)t + \phi_2]\} \\ &+ 0.0408 \{\sin[(\omega + 3f)t - \phi_3] - \sin[(\omega - 3f)t + \phi_3]\}. \quad (23) \end{aligned}$$

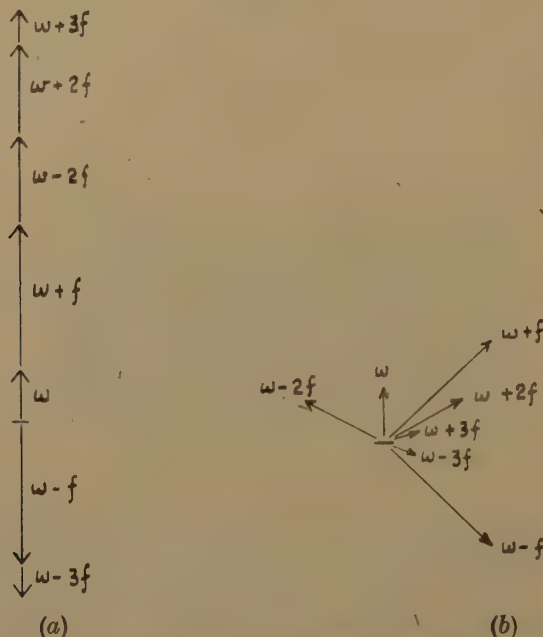
The vector diagram for carrier and side-bands at $t=0$ is shown in fig. 8 (a) for the original wave, and in fig. 8 (b) for the wave after passing through the tuned circuit. The terms of equation (23) must now be resolved into components in phase with the carrier and components in quadrature:—

$$\left. \begin{aligned} E_0 &= 0.2239 + 2 \times 0.1987 \cos(2ft - \phi_2). \\ E_q &= 2 \times 0.4074 \sin(ft - \phi_1) + 2 \times 0.0408 \sin(3ft - \phi_3). \end{aligned} \right\} \quad (24)$$

The phase of the resultant with respect to the carrier phase (*i. e.*, the phase modulation) is then

$$\phi_R = \tan^{-1}(E_q/E_0)$$

Fig. 8.



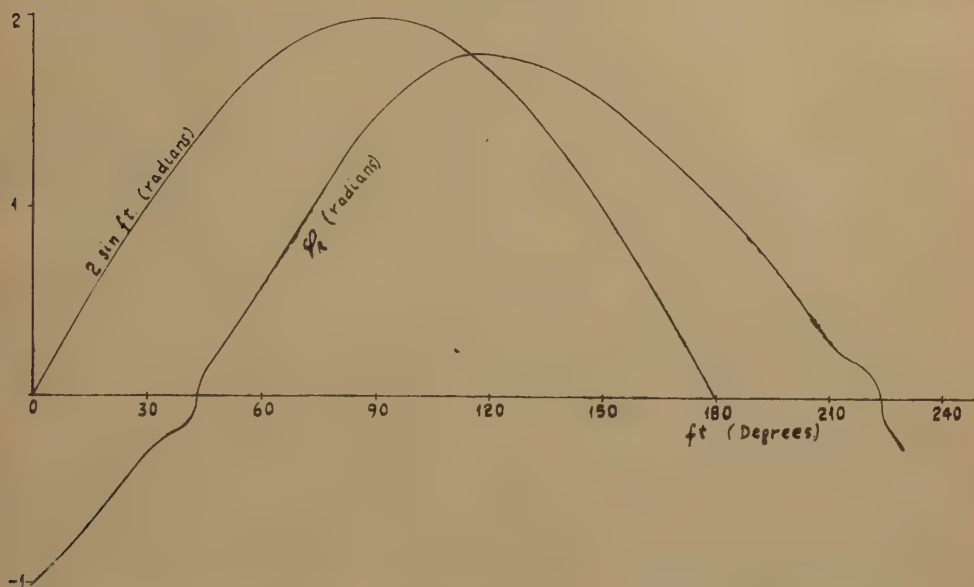
- (a) Relative phases of carrier and side-bands in F.M. sinusoidally-modulated signal.
 (b) Relative phases of carrier and side-bands (a) after passing through selective tuned circuit.

as compared with $2 \sin ft$ for the original wave of equation (22). This has been computed for a half-cycle of modulation, and is plotted in fig. 9.

This is the phase modulation, and the frequency modulation is derived from it by differentiation; this has been done graphically, and the resulting frequency modulation plotted in fig. 10. This shows very severe distortion, and, in fact, is similar to the wave-form demonstrated by Crosby in connection with the effects of selective fading⁽⁶⁾; it must, of course, be realized that the more obvious distortion of wave-form

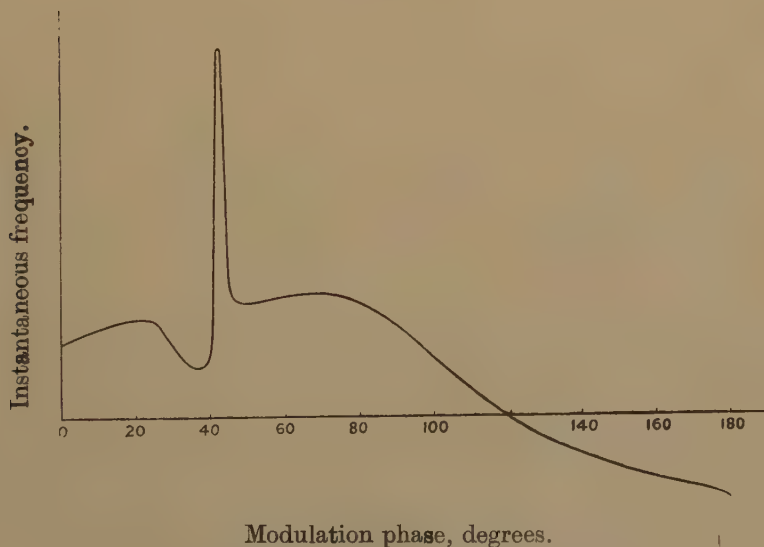
consists of high-order harmonics of the original modulation, which may then be neglected if they fall outside the frequency-range of the remainder of the communication system.

Fig. 9.



Distorted phase-modulation resulting from application of sinusoidal F.M. of wide frequency sweep to narrow circuit.

Fig. 10.



Modulation phase, degrees.
Distorted frequency modulation corresponding to the distorted phase modulation of fig. 9.

The applied signal was taken as $\sin(\omega t + 2 \sin ft)$, so that its frequency modulation was $2f \cos ft$; the retardation of phase produced by passage through the tuned circuit is evident in fig. 10, where the fundamental component does not reach its maximum until some time after $ft=0$.

Considering the difference between the square and sinusoidal modulating wave-form, there seems to be a good parallel between the dip and peak in the distorted wave of fig. 10 and the shoulder A and peak B in fig. 5.

8. CONCLUSIONS.

The side-band analysis of the response of a resonant circuit to a simple form of sinusoidally frequency-modulated signal gives indirect confirmation of the results of the operational method of treating square-wave frequency modulation developed in this paper. It should be noted that the square-wave results were calculated for changes of frequency symmetrical about the resonant frequency of the circuit in question; owing to the complexity of the expression for an asymmetrical change, *e. g.*, the infinitesimal change about an arbitrary frequency of resonance, there seems little hope of developing an equivalent of Heaviside's superposition theorem to relate the square-wave F.M. response to the response to frequency modulation of sinusoidal or arbitrary wave-form. For a given ratio of tuned circuit band-width to maximum frequency accepted by the modulation-frequency portions of the system, a wide-deviation frequency modulation system gives a better transient response than an amplitude-modulation system.

The only circuit whose response has been examined is the simple L. C. R resonant circuit; the behaviour of more complex circuits (*e. g.*, "band-pass" transformers) may either be inferred qualitatively from the single-circuit results, determined empirically, or calculated for any given case.

9. ACKNOWLEDGEMENTS.

The author wishes to thank Mr. L. H. Bedford for his encouragement and interest in this work, and the management of A. C. Cossor, Ltd., for permission to publish it.

References.

- (1) Salinger, "Transients in Frequency-Modulation," *Proc. I. R. E.* xxx. p. 378 (1942).
- (2) Bell, "Theory of Ideal Filters," *Wireless Engineer*, xx. p. 323 (1943).
- (3) Bell, "Reduction of Band-width in F.M. Receivers," *Wireless Engineer*, xix. p. 497 (1942).
- (4) Bell, "Characteristics of F.M. Systems," *Wireless Engineer*, xx. p. 233 (1943).
- (5) Van der Pol, "Frequency Modulation," *Proc. I. R. E.* xviii. p. 1194 (1930).
- (6) Crosby, "F.M. Propagation on 26 Mc/s.," *Proc. I. R. E.* xxix. p. 398 (1941).

XVII. *Graphical Computation.*

By C. W. HANSEL, Bedford School *.

[Received March 11, 1943.]

[The original paper had to be curtailed to half its length owing to paper shortage and other difficulties. The matter eliminated included the computation of $\Sigma(a^2)$ and $\Sigma(a^3)$ by three different methods, and a discussion of graphical computations with the aid of several charts in combination or book charts.]

1. *Introduction.*

It is not so long ago that seven-figure mathematical tables were used by students, engineers and scientists, in computations in which three or four-figure accuracy would have been optimistic. Four-figure tables are now in general use and practical men often use slide rules, but the vagueness with regard to the degree of accuracy of data and results remains. Most students work consistently to four figures without considering whether this degree of accuracy is justified or not.

Sir C. G. Darwin, in addressing a joint meeting of the Civil, Electrical and Mechanical Engineers (April 1942), made the following significant remarks: "I concluded that in the case of a good many engineers there was a defect in the habits of thought which they had been taught. I would not like to say that that applied only to engineers; nearly all education in this and many other countries had suffered from the same sort of thing until comparatively recently. People were taught to think of a dimension or quantity as an exact number or magnitude, whereas the proper way to think about every dimension was to regard it as having a fringe, as being a number plus or minus a little bit, and the magnitude of that little bit was a very important quality of the number."

A glance at any examination paper or textbook of mathematics, science or engineering reveals how unsatisfactory are the accepted standards by which magnitudes are habitually expressed. Looking closer still, how easily a vague slovenly statement of quantity trips from the tongue or the pen. The great advantage of the graphical representation of data is that it reveals conspicuously and at a glance deviations and characteristics of the data which can only be obtained laboriously and piecemeal from a table of figures. Moreover, the accuracy of the graphical method can be made appropriate to the accuracy of the data.

In the discussion following a paper by R. C. Strachan on "Nomographic Solutions for Formulas of Various Types" (Trans. Amer. Soc. of

* Communicated by the Author.

Civ. Eng. lxxviii. (1915) pp. 1359-1408), Allen Hazen said : " Engineers, like other people, have strongly fixed habits. Some engineers can best solve their problems on a slide rule, some by the use of diagrams, and others by printed tables. For the solution of the Williams-Hazen hydraulic formula, at least five procedures may be used : (1) the hydraulic slide-rule made by Ledder and Probst. Boston ; (2) ' Hydraulic Tables ' published by John Wiley and Sons ; (3) diagrams on logarithmic paper, of which there are a number of very excellent ones, such as those published in some of the engineering papers ; (4) the procedure suggested by the author ; and (5) direct solution with a log-log slide rule."

The next speaker, Carl B. Andrews, said : " It seems . . . that many of the mathematical operations which occur in civil engineering work require a higher degree of accuracy than is obtainable from diagrams of any kind, and that the diagrams find their proper use in checking computations and in making quick approximate calculation. . . . A classroom remark of the late Professor Thomas Gray, which strongly impressed the writer, was to the effect that there was a disadvantage in the use of reckoning machines, diagrams, multiplication tables, and other mechanical aids in computation, in that the computer was likely to get into the habit of writing down the result without noting the intermediate steps, which made it very hard to trace errors resulting from an accidental wrong entry or wrong reading. In the preparation of the Smithsonian Physical Tables, Dr. Gray said that much computation had been required, but that comparatively few errors had been found after its publication, because the computation had been carried out in full in note-books, where they were readily checked."

The writer is not in agreement with either of the opinions here expressed. The accuracy of graphical charts can be made to exceed that of printed tables page for page. Secondly, any mistake is at once obvious on a graphical chart—there is no need to search for it or to go over back work to look for antecedent errors. The correction is equally obvious.

2. *Graphical Tables.*

Gunter was the first to use a functional scale for purposes of computation. He used a logarithmic scale and dividers to perform multiplication and division. Quoting from " Calcul Graphique et Nomographie " (par M. D'Ocagne), footnote, p 168, " L'échelle logarithmique, imaginée en 1624 par Gunter, et que Wingate appliquait la même année à la construction de la règle à calcul, est sans doute le premier exemple connu d'une échelle fonctionnelle."

Gunter's scale has been dated 1620, and Wingate's rule 1628. Oughtred's slide rule 1630 (' Handbook of Napier Centenary,' p. 155, Dr. G. D. C. Stokes). Mannheim added the cursor in 1851.

M. J. Eichhorn, ' Western Electrician,' March 9, 1901, p. 162. " The Construction and Use of Graphical Tables," says : " Dumesnil, in Paris, issued in 1894 a magnified edition of Gunther's scale, and this is a striking instance of the superiority of graphical representation for certain purposes.

Dumesnil's scale is 20 metres, or 65 feet, long, divided in 80 pieces, which are placed side by side, and occupy only 4 pages. It is equivalent to an ordinary five-place logarithm table occupying 30 pages. . . .

"The system of graphical tables briefly outlined above opens up great possibilities of practical application in all branches of engineering, and ultimately all numerical calculations required in practical work may be reduced to reading the results on graphical tables prepared in advance by means of dividing engines, and providing for every possible combination of circumstances".

3. Notation.

Scales may be lettered or numbered, the letter or number being enclosed by a square bracket. For example, [A] reads "scale A" and [1] reads "scale 1."

If "P is a point on scale A where the reading is a ," this may be written $P=a[A]$, and reads "P is a on A."

If $P=a[A]$ and $Q=b[B]$ and the join PQ passes through $R=c[C]$, this may be denoted by $a[A]:b[B]\rightarrow c[C]$. This may be read "A, B, C, are in alignment."

The arrow usually signifies more than mere alignment. It usually implies that some mathematical operation involving a and b gives a result c .

The line used to obtain alignment is called a "secor."

The scales A and B are primary scales, and the scale on which the result is read, namely C, is a secondary scale.

If a normal to a scale at a point $P=a[A]$ passes through $b[B]$, a and b are corresponding readings, and this may be indicated by " $a[A]\equiv b[B]$ " and reads " a on A is opposite b on B."

A long scale may be broken into shorter lengths or subscales. Subscales are usually arranged with their axes parallel to one another. The subscales of scale A are usually denoted by A_0, A_1, A_2 , etc., or, if s and t are integers, by A_s or A_t .

If a reading $a[A]$ is to be transferred to the same reading on scale B, this is denoted by $a[A] \nearrow a[B]$.

4. Graphical Charts.

Fig. 1 illustrates one of ten graphical charts numbered 0 to 9. The chart number is at the lower left-hand corner under A, and fig. 1 illustrates, in outline only, the arrangement of the scales on chart number 6, or pages 6 and 6-5. A pile of 10 charts may be used in association, or each chart may be arranged as in fig. 1, so as to form two pages of a book of 20 pages in all. The first page and the first chart are numbered 0. The second page is 0-5. The oblique scales on the right half of chart number 6 do not belong to this chart.

Each chart is graduated to general scales A and A'. The tabular scales B, C, C' D, D', are graduated and figured to the left of the subscales and the scales B', C'', D'', are graduated and figured to the right of the subscales.

The figuring of the general scale A is with reference to the chart number below A and to the numbers in the horizontal and vertical margins opposite A. For example, if a is the reading on scale A, corresponding to 42 on subscale B₂, $a=62325$, the digits being in succession the chart number 6, followed by the subscale number 2, followed by the reading on the uniform vertical scale 325.

In fig. 1, $\log b=a$ and $\text{illog } a=b$. Omitting decimals and characteristics, $\log 42000=62325$, and $\text{illog } 62325=42000$.

If b_0, c_0, d_0 ; b_1, c_1, d_1 ; etc., are corresponding readings on subscales B₀, C₀, D₀; B₁, C₁, D₁, etc., scales C and D are graduated in degrees so that

$$\begin{aligned}\sin c &= b \text{ and therefore } \log (\sin c) = a, \\ \tan d &= b \text{ and therefore } \log (\tan d) = a.\end{aligned}$$

Chart 6 covers a range of a from 60000 to 70000.

Subscales C'' and D'' are graduated in radians.

$$c \text{ degrees} = c'' \text{ radians}; \quad d \text{ degrees} = d'' \text{ radians}.$$

$$\begin{aligned}\text{Also,} \quad \sin c'' &= b; \quad \log (\sin c'') = a, \\ \tan d'' &= b; \quad \log (\tan d'') = a.\end{aligned}$$

For example, using subscales C, C'', $23^\circ 42'$ or $23.700^\circ = 0.41364^\circ$; $\sin 0.41364 = 0.40195$; $\log (\sin 0.41364) = 0.60417$.

Again, using subscales D, D'', $\tan 21^\circ 53' 50'' = \tan 0.38218 = 0.40195$; $\log (\tan 21^\circ 53' 50'') = 0.60417$.

The general scale A' and the tabular scales B', C', D', are negatively figured for backward reading, the figuring of these scales being marked below the graduation mark. Red figuring also serves to distinguish backward readings.

Scale B' is a reciprocal scale— b and b' are reciprocals.

$$\begin{aligned}\text{Hence, since} \quad \sin c &= b; \quad b' = 1/b = \text{cosec } c. \\ \text{also} \quad \tan d &= b; \quad b' = \cot d.\end{aligned}$$

The graduation marks of scales C' and D' coincide with those of scales C and D, but scales C' and D' are figured for backward reading and below the graduation lines and preferably in red.

$$c + c' = 90^\circ \quad \text{and} \quad d + d' = 90^\circ.$$

$$\begin{aligned}\text{Hence, } \cos c' &= \sin c = b; \quad \log (\cos c') = a; \quad b' = \sec c'; \quad \cot d' = \tan d = b; \\ &\log (\cot d') = a.\end{aligned}$$

Ten charts similar to that of fig. 1 give direct and inverse readings of logs, reciprocals, trigonometrical functions and their logarithms for both radians and degrees, to five figure accuracy. The total area of the 10 charts is about 10,000 square centimetres. Direct and inverse readings are obtained with equal facility and without interpolation. Such a graphical table is much more compact than mathematical tables and not so difficult to use.

If a higher degree of accuracy is required, more than 10 charts may be used in association.

5. *Scale Co-ordinates.*

The right half of fig. 1 has oblique scales added to illustrate another method of arranging scales. The oblique scales do not belong to this chart, but have been drawn on fig. 1 to avoid duplication of printing blocks. A special arrangement of scales may be desirable for the tabulation of certain data, for example, steam tables, wire tables, magnetic properties of materials etc.

In fig. 1 the scales E, F, G are each graduated according to a different method.

Scale E is graduated so that $\log (\sin e)=a$; for example,

$$P=\log (\sin 26.800^{\circ})=1.65406.$$

Scale F has a scale co-ordinate 42° marked near each end of it. Scale F is graduated so that $\sin f=a$, for example,

$$Q=\sin 42.300^{\circ}=0.67301.$$

Scale G has a scale co-ordinate of 93, which figures precede the reading on the vertical A subscales. Scale G is figured with respect to the horizontal marginal numbers and is graduated so that :

$$\sin g=0.93+\text{reading on vertical A subscale} \times 10^{-2};$$

for example, $R=\sin 68.500^{\circ}=0.93+0.042 \times 10^{-2}=0.93042$.

6. *Graphical Computation.*

Compactness is not the only advantage of a graphical table. Graphical charts not only give tabular information without interpolation, but they may be used for the actual computation.

In fig. 1, let x, y be whole number readings on scale A.

Let x occur on chart number p , subscale number s , and at reading a_s on the subscale. This may be written $x[A]=p, s, a_s$.

Similarly, let y occur on chart q , subscale t , and at subscale reading a_t , or $y[A]=q, t, a_t$.

Then $z=(x+y)$ occurs on chart $(p+q)$, subscale $(s+t)$, and at subscale reading (a_s+a_t) .

$$\begin{aligned} \text{Thus} \quad z &= (x+y) = (p+q), (s+t), (a_s+a_t), \\ \text{and} \quad z &= 10^4(p+q) + 10^3(s+t) + 10^2(a_s+a_t). \end{aligned}$$

For example, if $x=12345$ and $y=53426$, $x=1, 2, 3.45$ and $y=5, 3, 4.26$, hence $z=6, 5, 7.71$, or $z=65771$ (ignoring decimal points).

Let X, Y, Z be readings on subscales B_s, B_t, B_{s+t} , corresponding to readings x, y, z on scale A.

$$\text{Then} \quad x=\log X, \quad y=\log Y, \quad z=(x+y)=\log XY=\log Z.$$

Thus, $Z=XY$. Hence, to find the product of two numbers X and Y , proceed as follows :—

$$\text{Find} \quad x[A] \equiv X[B], \text{ and let } x[A]=p, s, a_s.$$

$$\text{Find} \quad y[A] \equiv Y[B], \text{ and let } y[B]=q, t, a_t.$$

Then, $z = 10^4(p+q) + 10^3(s+t) + 10^2(a_s + a_t)$.

Also, $z[A] = Z$ or $XY[B]$.

The following rules are more easily invented than followed :—

If $(p+q) < 10$, there are five digits in the result, the first being $(p+q)$, otherwise there are six, the first two being $(p+q)$.

If $(s+t) < 10$, the thousands digit is $(s+t)$, otherwise add one to the ten-thousands digit and use the units digit of $(s+t)$ as the thousands digit of the result. The remaining three digits are $10^3(a_s + a_t)$. If this exceeds 999 add one to the thousands digit.

If $Z = X/Y$, Z may be found by multiplying X by the reciprocal of Y . Obviously scale B' is used to give the reciprocal of Y . $Y[B'] \equiv 1/Y[B]$.

If $Z = X^Y$, $\log Z = Y \log X$.

To find Z , first find $\log X$, $X[B] \equiv \log X[A]$, $\log X[A] \nearrow \log X[B]$. $\log X$ is now multiplied by Y , as already described, the result being $\log Z[A]$. Finally, $Z[B] \equiv \log Z[A]$.

It will be apparent that $X \pm Y$, XY , X/Y , XY may be found, using graphical charts, such as illustrated in fig. 1. Hence, x^n , $a \sin x$, $a \sin x + b \cos x$, etc., may be evaluated graphically.

The least and greatest values of the tabulated quantities are indicated on each chart. This facilitates the location of scale readings.

7. Computation of x^n .

Fig. 2 is a graphical table and nomogram combined. There are general scales A and B for forward reading, and A' and B' for backward reading. The subscales A and A' are duplicated and numbered 0, 0 to 35, 35. There are five A subscales along each axis numbered and figured marginally in the margins opposite A (ignoring the decimal point for figuring). Similarly, there are five A' subscales along the same axes figured for backward reading in the margins opposite A' . The B and B' subscales are numbered 0 to 70, with 10 subscales along each axis. They are numbered and figured in the margins B and B' .

Scales C and subscales C are lolog scales graduated lologarithmically to scale A and figured to the left of the axis.

$$\text{lolog } c = a; \quad c = \text{lolog}^{-1}a \text{ or } \text{illolog } a; \quad -\text{lolog } (1/c) = a'.$$

Scale C' and subscales C' are reciprocal scales figured to the right of the axes. c and c' are reciprocals.

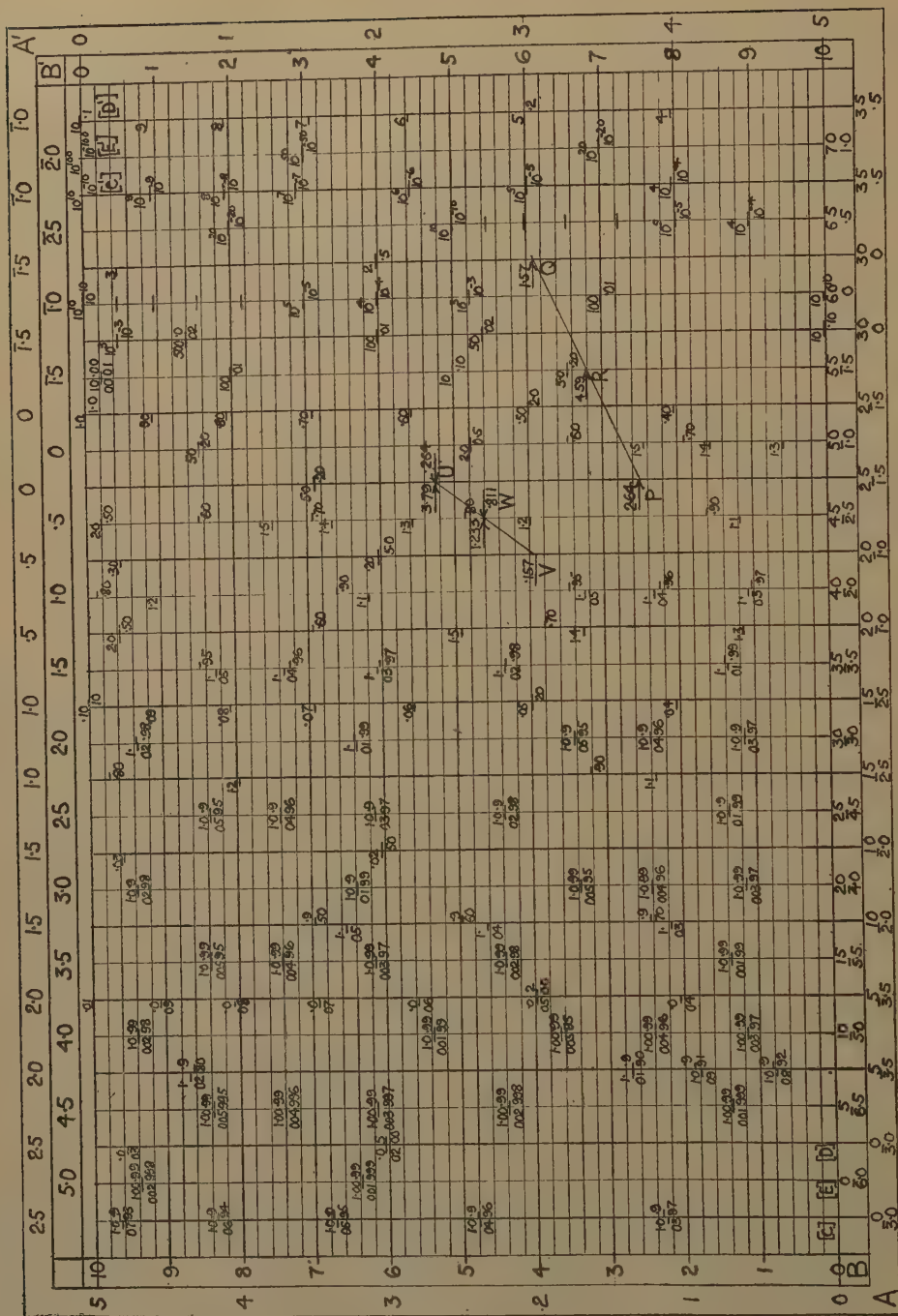
$$\text{lolog } c' = a'; \quad c' = \text{illolog } a'; \quad -\text{lolog } (1/c') = a.$$

Scales and subscales D and D' are graduated logarithmically to scales A and A' . Scale D is figured forwards on the left of the axis, and scale D' is figured backwards on the right of the axis. d and d' are reciprocals.

$$\log d = a; \quad d = \log^{-1}a \text{ or } \text{illog } a.$$

$$\log d' = a'; \quad d' = \log^{-1}a' \text{ or } \text{illog } a'.$$

Fig. 2.



Subscales C and D, C' and D' are numbered in the same way, but the C and C' subscales are in each case to the left of the D and D' subscales of the same number.

Scales and subscales E and E' are in groups of 10 along each axis and graduated lologarithmically to scales B and B'. E subscales are figured forwards to the left and E' subscales are figured backwards on the right of the axes. e and e' are reciprocals.

$$\text{lolog } e = a; \quad e = \text{illolog } a.$$

$$\text{lolog } e' = a'; \quad e' = \text{illolog } a'.$$

If $y = x^n$ and $y' = x^{-n}$,

$$x[C_s] : n[D_t] \rightarrow y[E_{s+t}] = y'[E'_{s+t}],$$

x has a range from 1.003 to 10^{10} ,

n has a range from 0.001 to 10, but the range of this scale may be increased indefinitely.

y has a range from about 1.00001 to 10^{100} .

If $x < 1$, $(1/x)^n = 1/y$. In this way, $y = x^n$ may be evaluated for values of x less than unity.

Examples :

1. To find $y = 2.64^{1.57}$.

$$P = 2.64[C_{25}] : Q = 1.57[D_{30}] \rightarrow R = 4.59[E_{55}].$$

$$\underline{y = 4.59.}$$

2. To find $y = 0.264^{0.157}$.

$$U = 0.264 = 1/3.79; \quad V = 0.157;$$

$$3.79[C_{25}] : 0.157[D_{20}] \rightarrow 1.233[E_{45}] \text{ or } 0.811[E'_{45}]$$

$$y = 1/3.79^{0.157} = 1/1.233 = 0.811.$$

3. To find $\text{lolog } 1.1$ and $\text{cololog } 0.90$.

$$1.1[C_{15}] = \bar{2}.617[A]; \quad \text{lolog } 1.1 = \bar{2}.617.$$

$$0.90[C_{15}] = \bar{2}.660[A]; \quad -\text{lolog } 0.90 = \bar{2}.660.$$

$$0.90[C'_{15}] = 1.340[A']; \quad \text{cololog } 0.90 = 1.340.$$

8. An Appeal.

In the present paper an attempt has been made to show that graphical tables are, page for page, more accurate than mathematical tables. They are more convenient and have a wider range of usefulness than mathematical tables, because readings are made without interpolation, and computations may be carried out on the charts themselves without transference of readings or subsidiary arithmetical work. Only skeleton scales have been drawn and rulings are not shown in detail. Graphical computation may be carried out using a book of charts in much the

same way as with a slide rule, but graphical charts are much more versatile and their accuracy is unlimited.

The author has discussed nomographic and other graphical methods in the following articles, either already published or to be published in the 'Philosophical Magazine':—

"Standard Curves";

"Graphical Differentiation and Integration."

Also, in the 'Mathematical Gazette':—

"The Graphical Treatment of Quadratic Equations."

In the opinion of the author, pure and applied mathematics, physics, mechanical and electrical engineering, etc., should all be combined into one subject up to the standard of a pass degree. Geometry should include mensuration, trigonometry, analytical geometry, solid geometrical drawing, machine drawing, perspective drawing—indeed, every form of geometrical representation. Heat engines and electrical engineering are parts of applied physics or mathematics, as are also applied mechanics, strength of materials and structures, hydraulics, theory of machines, etc. Graphical and analytical methods should be used either alternatively or jointly.

The object of all the previously mentioned papers is to indicate that the neglect of so useful an aid to scientific and engineering work as nomography, is due to lack of facilities and apparatus for using nomographic methods. Before logarithmic tables were published or calculating machines were used, calculations were carried out laboriously with prodigious waste of time. Before squared paper could be bought cheaply it was scarcely used by anyone. If accurate nomographic paper (that is paper with functional scales printed on the paper ready for use) were available as cheaply as squared paper, it could, with advantage, be used even more extensively. Instead of this, the subject of nomography is almost unknown in schools and colleges, and engineers who realize its value waste a great deal of time in drawing accurate nomograms. Far more time must be spent in drawing a nomogram than in drawing a sheet of squared paper. Not only this, but nomography remains almost stagnant in development. Little has been added to the subject since D'Ocagne published his "*Traité de Nomographie*" (1899) and "*Calcul Graphique*" except an article on "Nomograms" by Joergen Rybner specially written and translated at the request of the General Electric Company of America and which appeared in the 'General Electric Review,' xxxiii, 1930.

A child of twelve can use a simple nomogram with very little explanation. Only an accurate and careful draughtsman can prepare a nomogram sufficiently accurate for useful calculation. Is there no lover of science or of engineering who is willing to provide the facilities for studying and developing the subject so that it may find a place in the school curriculum? If so, there is both opportunity and privilege. Extensive research is desirable in devising and perfecting machines for ruling and

graduation, for investigating printing processes and machinery capable of giving accurate register of rulings, graduations and numbers in different colours. Also, research on paper, celluloid, cellophane and other bases on which nomographic scales, curves, etc., can be printed.

What the slide rule achieves for three-figure accuracy, that single sheets of nomographic paper can achieve at much less cost and with greater versatility and accuracy. What the five- or six-figure mathematical table provides for more accurate calculations, that and more a book of mathematical charts can achieve with greater thoroughness than any printed tables that are available.

There are many possibilities of useful application of graphical charts to tabulation of data and computation, such as steam tables, magnetic data, hygrometry, spectrographic data, navigation and astronomy, etc. These will only be developed and used when graphical methods and charts are used by schoolboys, artisans, and shopkeepers.

The following are some examples of the application of graphical methods in industry :—

Prof. Reginald O. Kapp, "Nomograms in Electrical Engineering." Journ. Inst. E. E. lxxviii. 1936, which includes a chart for a two-part tariff.

E. L. McMillen, "Versatile Nomograph for Chemical Engineering Calculations," Ind. and Eng. Chem. xxx. 1938.

W. M. Armstrong, "Nomogram for Routine Illumination Computation," Illum. Eng. xxxvi. 1941.

R. L. Peek, Jnr., "Universal Alignment Chart," 'Bell Laboratories Record,' xiii. 1935.

H. G. Weist, Jnr., "Advantage of Logarithmic Graphs in Design Engineering," 'General Electric Review,' xli. 1938.

Designers should realize that here is a means of saving time and effort which should not be neglected. A large amount of information which has been kept in the form of tables and more or less unsatisfactory rectangular graphs may be preserved in more convenient and accurate form by reducing it to logarithmic graphs.

XVIII. Some Problems involving Line Sources in Conduction of Heat.

By J. C. JAEGER *.

[Received February 12, 1943.]

1. Introduction.

THE object of this paper is to give the solutions of two problems in Conduction of Heat which are particularly easily solved in cylindrical co-ordinates by the Laplace transformation method because of the

* Communicated by the Author.

special form [(14), (15) below] of the Laplace transform of the temperature due to a line source.

The first of these is the temperature due to a line source, parallel to the z -axis, in a composite solid for which the region $r < a$ is of one substance and $r > a$ of another. Problems involving line sources in composite medium do not appear to have been studied hitherto and the results are of some practical importance.

The second problem is that of the temperature of a circular cylinder round whose surface a line source moves with constant angular velocity ω . This case arises in practice when a rotating cylinder is frictionally heated by rubbing portion of its surface. More complicated cases in which some portions of the surface are heated by friction in this way while other portions are kept at constant temperature are also discussed.

2. To find the temperature due to a unit instantaneous line source at zero time, parallel to the z -axis, through the point whose cylindrical co-ordinates are (r', θ') in composite solid, for which the region $0 \leq r < a$ has conductivity K_1 , density ρ_1 , specific heat c_1 , and diffusivity $\kappa_1 = K_1/\rho_1 c_1$, and in the region $r > a$, the corresponding quantities are K_2 , ρ_2 , c_2 , κ_2 .

We consider the case $r' < a$, that of $r' > a$ is treated similarly.

The temperature at time t at (r, θ) , due to an instantaneous unit line source at zero time at (r', θ') in infinite medium of constants K_1 , κ_1 , is given * by

$$u = \frac{1}{4\pi\kappa_1 t} e^{-R^2/4\kappa_1 t}, \quad \dots \dots \dots (1)$$

where $R^2 = r^2 + r'^2 - 2rr' \cos(\theta - \theta'). \quad \dots \dots \dots (2)$

We seek a solution for the temperature v_1 in $0 \leq r < a$ of the form

$$v_1 = u + w, \quad \dots \dots \dots (3)$$

where w has to satisfy the equation of conduction of heat in this region, and to tend to zero as $t \rightarrow 0$. If v_2 is the temperature in the region $r > a$ it has to satisfy the equation of conduction of heat in this region, and to tend to zero as $t \rightarrow 0$, and as $r \rightarrow \infty$. The boundary conditions to be satisfied at $r = a$ are

$$v_1 = v_2, \quad \dots \dots \dots (4)$$

$$K_1 \frac{\partial v_1}{\partial r} = K_2 \frac{\partial v_2}{\partial r}. \quad \dots \dots \dots (5)$$

Let the Laplace transforms † of u , w , v_2 be \bar{u} , \bar{w} , \bar{v}_2 , i. e.

$$\bar{u} = \int_0^\infty e^{-pt} u(t) dt, \quad p > 0, \quad \dots \dots \dots (6)$$

* Carslaw, 'Conduction of Heat,' Ed. 2, § 72 (1921). The quantity of heat liberated per unit length is $\rho_1 c_1$.

† For the application of the Laplace transformation method to problems involving line sources in Conduction of Heat see Carslaw and Jaeger, Phil. Mag. (7) xxxi. p. 204 (1941).

etc. Then \bar{w} has to satisfy

$$\frac{\partial^2 \bar{w}}{\partial r^2} + \frac{1}{r} \frac{\partial \bar{w}}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \bar{w}}{\partial \theta^2} - q_1^2 \bar{w} = 0, \quad 0 \leq r < a, \quad (7)$$

where

$$q_1 = \sqrt{(p/\kappa_1)}. \quad (8)$$

And \bar{v}_2 has to satisfy

$$\frac{\partial^2 \bar{v}_2}{\partial r^2} + \frac{1}{r} \frac{\partial \bar{v}_2}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \bar{v}_2}{\partial \theta^2} - q_2^2 \bar{v}_2 = 0, \quad r > a, \quad (9)$$

where

$$q_2 = \sqrt{(p/\kappa_2)}. \quad (10)$$

Also at $r=a$ we require

$$\bar{u} + \bar{w} = \bar{v}_2, \quad (11)$$

$$K_1 \frac{d\bar{u}}{dr} + K_1 \frac{d\bar{w}}{dr} = K_2 \frac{d\bar{v}_2}{dr}. \quad (12)$$

It is known * that

$$\bar{u} = \frac{1}{2\pi\kappa_1} K_0(q_1 R), \quad (13)$$

$$= \frac{1}{2\pi\kappa_1} \sum_{n=-\infty}^{\infty} \cos n(\theta - \theta') I_n(q_1 r) K_n(q_1 r'), \quad r < r', \quad (14)$$

$$= \frac{1}{2\pi\kappa_1} \sum_{n=-\infty}^{\infty} \cos n(\theta - \theta') K_n(q_1 r) I_n(q_1 r'), \quad r > r'. \quad (15)$$

From these equations and the conditions that \bar{w} is to remain finite as $r \rightarrow 0$, and \bar{v}_2 as $r \rightarrow \infty$, we obtain

$$\bar{v}_2 = \frac{K_1}{2\pi\kappa_1} \sum_{n=-\infty}^{\infty} \frac{1}{\Delta} I_n(q_1 r') K_n(q_2 r) \cos n(\theta - \theta'), \quad (16)$$

$$\bar{v}_1 = \frac{1}{2\pi\kappa_1} \sum_{n=-\infty}^{\infty} \frac{1}{\Delta} I_n(q_1 r') [\Delta K_n(q_1 r') - D I_n(q_1 r')] \cos n(\theta - \theta'), \quad 0 < r < r', \quad (17)$$

where $\Delta = K_1 q_1 I'_n(a q_1) K_n(a q_2) - K_2 q_2 I_n(a q_1) K'_n(a q_2)$, (18)

and $D = K_1 q_1 K'_n(a q_1) K_n(a q_2) - K_2 q_2 K_n(a q_1) K'_n(a q_2)$, (19)

and to find the value of \bar{v}_1 when $r' < r < a$ we interchange r and r' in (17).

The values of v are obtained from those of \bar{v} by the inversion theorem for the Laplace transform, namely

$$v(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{st} \bar{v}(\lambda) d\lambda. \quad (20)$$

Thus, writing $\mu_1 = \sqrt{(\lambda/\kappa_1)}$, and $\mu_2 = \sqrt{(\lambda/\kappa_2)}$, we have

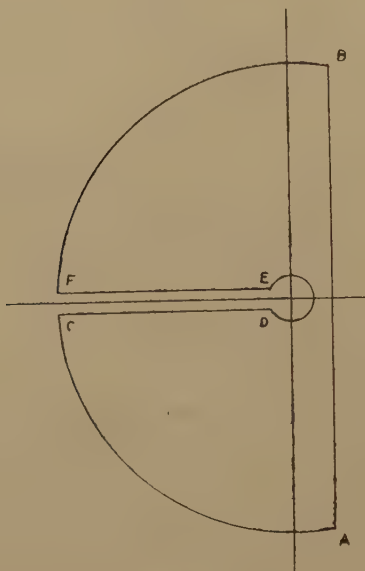
$$v_2 = \frac{K_1}{4\pi^2 i a \kappa_1} \sum_{n=-\infty}^{\infty} \cos n(\theta - \theta') \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{st} I_n(\mu_1 r') K_n(\mu_2 r) d\lambda}{\Delta(\lambda)}, \quad (21)$$

with an expression of the same type for v_1 .

* Gray and Mathews, 'Treatise on Bessel Functions' (Ed. 2), p. 51 (31) and p. 74 (59).

The line integrals in (21) are evaluated by considering the contour of fig. 1; it is easy* to show that there are no poles of the integrands in or on the contour, and that the integrals over the large circular arcs vanish in the limit as the radius tends to infinity. Thus, the line integrals

Fig. 1.



in (21) may be replaced in the limit by the integral over the path CDEF, and this gives a real infinite integral. The final results are

$$v_2 = \frac{K_1}{\pi^2 a} \sum_{n=-\infty}^{\infty} \cos n(\theta - \theta') \int_0^{\infty} \frac{e^{-\kappa_1 u^2} J_n(ur') [\psi J_n(\kappa ur) - \phi Y_n(\kappa ur)] du}{\phi^2 + \psi^2}, \quad (22)$$

$$v_1 = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \cos n(\theta - \theta') \int_0^{\infty} \frac{e^{-\kappa_2 u^2} J_n(ur) J_n(ur') [\phi g - \psi f] u du}{\phi^2 + \psi^2}, \quad (23)$$

where $\phi = \kappa K_2 J_n(au) J'_n(\kappa au) - K_1 J'_n(au) J_n(\kappa au), \quad (24)$

$$\psi = \kappa K_2 J_n(au) Y'_n(\kappa au) - K_1 J'_n(au) Y_n(\kappa au), \quad (25)$$

$$f = \kappa K_2 Y_n(au) J'_n(\kappa au) - K_1 Y'_n(au) J_n(\kappa au), \quad (26)$$

$$g = \kappa K_2 Y_n(au) Y'_n(\kappa au) - K_1 Y'_n(au) Y_n(\kappa au), \quad (27)$$

$$\kappa = \sqrt{(\kappa_1/\kappa_2)}. \quad (28)$$

3. A line source, emitting Q heat units per unit length per unit time, moves round the surface of a cylinder of radius a and thermal constants K, κ with uniform angular velocity ω . There is no loss of heat from the surface.

The Laplace transform of the temperature v at (r, θ) in a cylinder of radius a , with no loss of heat from the surface, due to a unit instantaneous

*. The work is elementary but tedious; for an example of the method see Jaeger, Proc. Roy. Soc. N.S.W. lxxv. p. 130 (1942).

line source at zero time at the point (a, θ') may be obtained from (17) by letting *

$$\kappa_1 = \kappa, \quad K_1 = K, \quad K_2 = 0, \quad r' = a,$$

which gives

$$\bar{v} = \frac{1}{2\pi\kappa a} \sum_{n=-\infty}^{\infty} \bar{v}_n \cos n(\theta - \theta'), \quad (29)$$

where

$$\bar{v}_n = \frac{I_n(qr)}{qI'_n(qa)}, \quad (30)$$

and

$$q = \sqrt{(p/\kappa)}. \quad (31)$$

From (29) we could, as in § 2, obtain v in the form

$$v = \frac{1}{2\pi\kappa a} \sum_{n=-\infty}^{\infty} v_n \cos n(\theta - \theta'), \quad (32)$$

though for the present purpose the v_n need not be evaluated explicitly.

Now suppose that, starting at zero time when the cylinder is at zero temperature, heat is emitted at a uniform rate of Q units per unit length per unit time along a line in the surface of the cylinder which starts from the point (a, β) and moves with uniform angular velocity so that its position at time t' is (a, θ') , where $\theta' = \omega t' + \beta$.

The temperature at time t at the point (r, θ) is, by (32),

$$V = \frac{Q}{2\pi a K} \int_0^t v_0(t-t') dt' + \frac{Q}{\pi a K} \sum_{n=1}^{\infty} \int_0^t \cos n(\theta - \omega t' - \beta) v_n(t-t') dt'. \quad (33)$$

The most convenient form for V is obtained by using the result that, if $\bar{x}_1(p)$ and $\bar{x}_2(p)$ are the Laplace transforms of $x_1(t)$ and $x_2(t)$, then $\bar{x}_1(p)\bar{x}_2(p)$ is the Laplace transform of

$$\int_0^t x_1(t-t') x_2(t') dt'. \quad (34)$$

Using (34), (30), and the result

$$\int_0^{\infty} e^{-pt} \cos(\omega t - \delta) dt = \frac{p \cos \delta + \omega \sin \delta}{p^2 + \omega^2}, \quad (35)$$

in (33) we obtain for the Laplace transform of V

$$\bar{V} = \frac{Q I_0(qr)}{2\pi a K p q I'_0(qa)} + \frac{Q}{\pi a K} \sum_{n=1}^{\infty} \frac{[p \cos n(\theta - \beta) + n\omega \sin n(\theta - \beta)] I_n(qr)}{(p^2 + n^2 \omega^2) q I'_n(qa)}. \quad (36)$$

We now obtain V from \bar{V} by applying the inversion theorem (20) to the terms of (36). The line integrals are evaluated by completing the contour by portion of a large circle, centre the origin, to the left of the imaginary axis, and evaluating the resulting contour integral as a sum of residues at its poles. The first term has a double pole at the origin, the terms of the series have simple poles at $\pm i n \omega$, and there are simple poles at $-\kappa \alpha_{ns}^2$, where $\pm \alpha_{ns}$, $s=1, 2, \dots$, are the roots of

$$J'_n(\alpha x) = 0. \quad (37)$$

* The result is easily obtained directly.

Evaluating the residues at these poles we have finally

$$V = \frac{2Q\kappa}{\pi K} \sum_{s=1}^{\infty} \sum_{n=1}^{\infty} \frac{\alpha_{n,s}^2 e^{-\kappa \alpha_{n,s}^2} J_n(r \alpha_{n,s}) [\kappa \alpha_{n,s}^2 \cos n(\theta - \beta) - n\omega \sin n(\theta - \beta)]}{(\kappa^2 \alpha_{n,s}^4 + n^2 \omega^2)(n^2 - \alpha^2 \alpha_{n,s}^2) J_n(a \alpha_{n,s})} \\ - \frac{Q}{\pi a^2 K} \sum_{s=1}^{\infty} \frac{e^{-\kappa \alpha_{0,s}^2} J_0(r \alpha_{0,s})}{\alpha_{0,s}^2 J_0(a \alpha_{0,s})} + \frac{\kappa Q t}{\pi a^2 K} \\ + \frac{Q}{2\pi a K} \sum_{n=1}^{\infty} \left\{ \frac{e^{in(\omega t - \theta + \beta)} I_n(i^{\frac{1}{2}} r \omega_n)}{i^{\frac{1}{2}} \omega_n I'_n(i^{\frac{1}{2}} a \omega_n)} + \text{conjugate} \right\}, \quad \dots \quad (38)$$

where $\omega_n = \sqrt{(n\omega/\kappa)}$. (39)

The first two terms of (38) are transient terms, the third is a non-periodic steady state term, and the last is the periodic steady state term*. If the motion has been going on for some time the transient terms will have disappeared, and in practice the increasing term $(\kappa Q t / \pi a^2 K)$, caused by accumulation of heat from the source, will usually be cancelled by some mechanism which removes heat from the solid. Thus we confine our attention to the periodic steady state terms. V_p , of (38). Using the notation

$$\text{ber}_n x + i \text{bei}_n x = i^n I_n(x i^{\frac{1}{2}}), \quad \dots \quad (40)$$

these terms become

$$V_p = \frac{Q\kappa^{\frac{1}{2}}}{\pi a K \omega^{\frac{1}{2}}} \sum_{n=1}^{\infty} n^{-\frac{1}{2}} M_n \cos \{n(\omega t - \theta + \beta) + \phi_1 - \phi_2\}, \quad \dots \quad (41)$$

where $\tan \phi_1 = \frac{\text{bei}_n r \omega_n}{\text{ber}_n r \omega_n}, \quad \tan \phi_2 = \frac{\text{bei}'_n a \omega_n}{\text{ber}'_n a \omega_n}, \quad \dots \quad (42)$

$$M_n = \left\{ \frac{\text{ber}_n^2 r \omega_n + \text{bei}_n^2 r \omega_n}{\text{ber}_n^2 a \omega_n + \text{bei}_n^2 a \omega_n} \right\}^{\frac{1}{2}} \dots \quad (43)$$

This result takes a specially simple form if $a\omega_1$ and $r\omega_1$ are large. In this case, using the asymptotic expansions of the Bessel functions, (41) reduces to

$$V_p = \frac{Q\kappa^{\frac{1}{2}}}{\pi K \omega^{\frac{1}{2}} r^{\frac{1}{2}} a^{\frac{1}{2}}} \sum_{n=1}^{\infty} n^{-\frac{1}{2}} e^{-(a-r)(n\omega/2\kappa)} \cos \{n(\omega t - \theta + \beta) - \frac{1}{4}\pi - (a-r)(n\omega/2\kappa)^{\frac{1}{2}}\}. \quad \dots \quad (44)$$

In the same way the problem of a line source rotating in the composite solid of § 2 can be solved. The most interesting case is that in which the source moves in the cylinder of separation, $r' = a$. In this case the temperature at the point (a, θ) in this cylinder is found, provided $a\sqrt{(\omega/\kappa_1)}$ and $a\sqrt{(\omega/\kappa_2)}$ are large, to be

$$\frac{Q(\kappa_1 \kappa_2)^{\frac{1}{2}}}{\pi a (K_2 \kappa_1^{\frac{1}{2}} + K_1 \kappa_2^{\frac{1}{2}}) \omega^{\frac{1}{2}}} \sum_{n=1}^{\infty} n^{-\frac{1}{2}} \cos \{n(\omega t - \theta + \beta) - \frac{1}{4}\pi\}, \quad \dots \quad (45)$$

* The advantage of the present method is that it separates out the physically interesting periodic terms in this way, which the solution obtained by direct evaluation of (33) does not,

which reduces to (44) if $K_2=0$. It may be remarked that the same factor,

$$\frac{(\kappa_1\kappa_2)^{\frac{1}{2}}}{K_2\kappa_1^{\frac{1}{2}}+K_1\kappa_2^{\frac{1}{2}}},$$

arises with an instantaneous plane source at the junction of two composite solids with a plane surface of separation*.

The problem of a cylinder of radius a with a rotating line source, and boundary condition

$$\frac{\partial v}{\partial r} + hv = 0,$$

can also be solved in the same way, but this boundary condition, though more general than the one discussed above, is not so useful as the latter in problems to be discussed subsequently in which the boundary condition varies round the surface of the cylinder.

4. A Rotating Source System in the Surface of a Cylinder.

If we put $Q=af(\beta)d\beta$ in (41) or (44) and integrate with respect to β from $-\alpha$ to α , we obtain the periodic steady state temperature at a point in a cylinder in whose surface heat is generated over an arc of angular width 2α , whose centre rotates with uniform angular velocity about the centre of the circle. The rate of heat generation is $f(\beta)$ per unit time per unit area at the point whose angular displacement from the centre of the arc is β . As before, there is no loss of heat from the surface. With the notation (42) and (43) we obtain from (41) for this temperature

$$\frac{\kappa^{\frac{1}{2}}}{\pi K \omega^{\frac{1}{2}}} \sum_{n=1}^{\infty} n^{-\frac{1}{2}} M_n \int_{-\alpha}^{\alpha} \cos \{n(\omega t - \theta + \beta) + \phi_1 - \phi_2\} f(\beta) d\beta. \quad (46)$$

The most important case is $f(\beta)=Q'$, a constant, which corresponds to the practical problem of generation of heat by friction from a moving contact of angular width 2α , the whole of the heat being supposed to be taken by the cylinder. In this case (46) becomes

$$\frac{2Q'\kappa^{\frac{1}{2}}}{\pi K \omega^{\frac{1}{2}}} \sum_{n=1}^{\infty} n^{-3/2} M_n \cos [n(\omega t - \theta) + \phi_1 - \phi_2] \sin n\alpha, \quad (47)$$

or, if $r\sqrt{(\omega/\kappa)}$ is large, using (44) we have

$$V_p = \frac{2Q'\kappa^{\frac{1}{2}}a^{\frac{1}{2}}}{\pi K \omega^{\frac{1}{2}}r^{\frac{1}{2}}} \sum_{n=1}^{\infty} n^{-3/2} e^{-(a-r)(n\omega/2\kappa)^{\frac{1}{2}}} \times \cos [n(\omega t - \theta) - \frac{1}{2}\pi - (a-r)(n\omega/2\kappa)^{\frac{1}{2}}] \sin n\alpha. \quad (48)$$

To study the way in which the temperature varies we take $\omega t=2m\pi$ in (47) or (48), so that the centre of the arc is on the initial line, and compute the result for various values of r and θ .

* Carslaw, *loc. cit.* § 77.

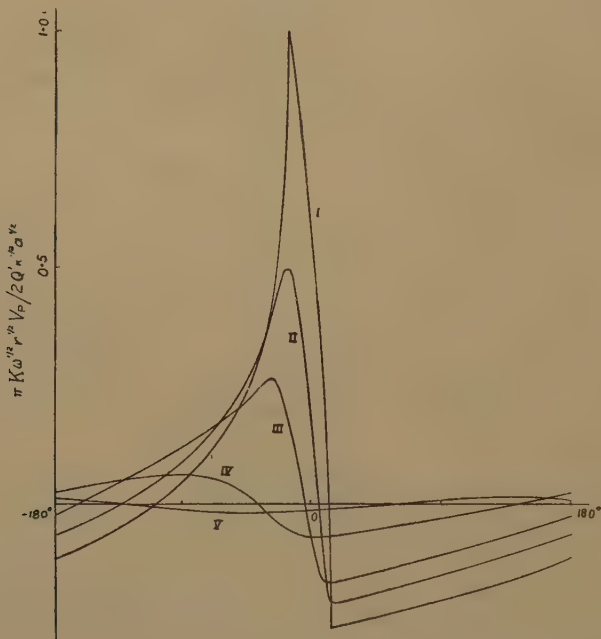
In fig. 2 results are shown for the case * (units are c.g.s. throughout)

$$\kappa=0.013, \quad \omega=8\pi, \quad a=78, \quad \alpha=15^\circ,$$

for which $r\sqrt{(\omega/\kappa)}$ is large, so (48) may be used, Curves I, . . . , V are the sums of the series in (48), i. e. the values of $(\pi K \omega^{\frac{1}{2}} r^{\frac{1}{2}} V_p / 2 Q' \kappa^{\frac{1}{2}} a^{\frac{1}{2}})$, for depths below the surface $(a-r)=0, 0.01, 0.02, 0.05, 0.1$ cm. respectively.

It is seen † that for the case computed these temperature oscillations die out very rapidly as we pass towards the interior of the cylinder, also that the maxima of the temperature curves are successively displaced to

Fig. 2.



the left : this effect being caused by the time taken for the heat generated on the surface to travel inwards.

5. Other Boundary Conditions.

The general problem of the type contemplated here is that of a rotating cylinder which is heated, for example by friction, over portion of its surface, while it is cooled over other portions of its surface, for example by dipping into liquid at constant temperature. The problem is thus one of a system of heat sources and sinks rotating round the

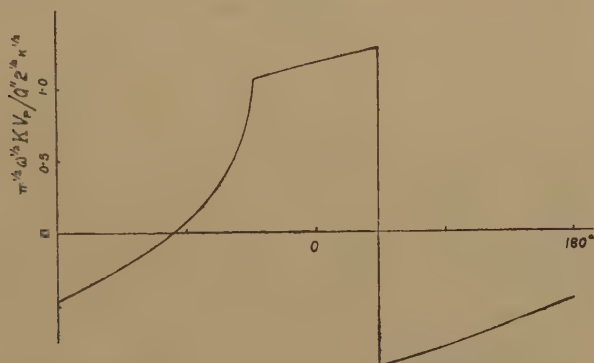
* The figures are for the sandstone cylinders used in the mechanical process of wood pulping : the problem will be referred to in § 6.

† For the corresponding results for a source moving in the plane surface of a semi-infinite region, see Jaeger, "Moving Sources of Heat and the Temperature of Sliding Contacts," Proc. Roy. Soc. N.S.W. lxxvi. p. 203 (1943).]

surface of the cylinder: in the steady state an amount of heat will be removed by the sinks equal to that supplied by the sources, and there will be periodic fluctuations of temperature given by (46).

The assumption of a constant source strength per unit time per unit area made in (47) is reasonable for regions in which heat is generated by friction, but for regions in which the surface is held at constant temperature the source strength $f(\beta)$ will not be constant. We can approach very nearly to this case by the following considerations: it is easily verified that, if the semi-infinite solid $x > 0$ is initially at zero and its surface $x = 0$ is to be maintained at constant temperature for $t > 0$, heat must be supplied at the origin at a rate proportional to $t^{-\frac{1}{2}}$; now in the case of an arc of width 2α rotating with uniform angular velocity round the surface of a cylinder, the point of the cylinder at angular distance $\beta < \alpha$ from the centre of the arc has been heated for a time $(\alpha - \beta)/\omega$, thus the above result for the semi-infinite solid suggests

Fig. 3.



that by taking $f(\beta) = Q'' (\alpha - \beta)^{-\frac{1}{2}}$ in (46), where Q'' is a constant, we should approximate to constant surface temperature over the arc of width 2α .

We consider only the surface temperature, *i. e.*, take $r = a$, and the case corresponding to (44) in which $a\sqrt{(\omega/\kappa)}$ is large. Then we obtain for the periodic steady state surface temperature

$$V_p = \frac{Q''}{K} \left(\frac{2\kappa}{\pi\omega} \right)^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{1}{n} \{ P(2n\alpha) \cos \delta_n + Q(2n\alpha) \sin \delta_n \}, \quad \dots (49)$$

where $\delta_n = n(\alpha - \theta) - \frac{1}{2}\pi + n\omega t, \quad \dots (50)$

and $P(x)$ and $Q(x)$ are Fresnel's integrals

$$P(x) = \frac{1}{2} \int_0^x J_{-\frac{1}{2}}(t) dt, \quad Q(x) = \frac{1}{2} \int_0^x J_{\frac{1}{2}}(t) dt. \quad \dots (51)$$

In fig. 3 the sum of the series in (49), *i. e.* the value of $\frac{KV_p}{Q''} \left(\frac{\pi\omega}{2\kappa} \right)^{\frac{1}{2}}$, is shown for $\alpha = .75$ radians. There is a discontinuity of $\pi/\sqrt{2}$ at $\theta = \alpha$,

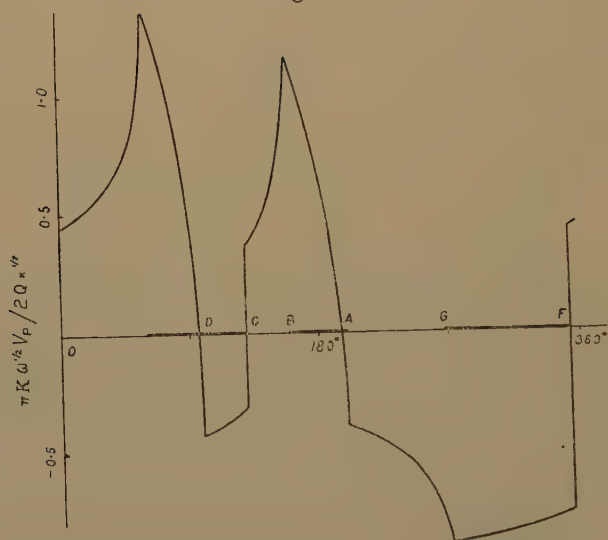
and the sum is seen to be nearly, though not accurately, constant for $-\alpha < \theta < \alpha$.

6. Any problem in which uniform heat sources or regions of constant temperature rotate with uniform angular velocity round the surface of a circular cylinder can be obtained by superposition of the solutions of

Fig. 4.



Fig. 5.



§§ 4 and 5 with appropriate phases. As an example we consider the system * of fig. 4.

* This approximates to that occurring in the mechanical system of wood pulping: wood is pressed on to a rotating grindstone in the regions AB and DE, the surface is cooled in the regions CD and FG, and there is negligible loss of heat in the regions BC, EF and GA. Some physical features of the practical problem are neglected here; a more elementary treatment which attempts to take these into account will be given elsewhere.

Over AB and DE there is uniform generation of heat, Q per unit time per unit area ; over CD and FG the temperatures are kept approximately constant, as in § 5, and such that $\frac{1}{4}$ of the heat generated in AB and DE is removed in CD, and $\frac{3}{4}$ in FG. There is no loss of heat in BC, EF, GA. It is assumed that $a\sqrt{(\omega/\kappa)}$ is large. Values of $\pi K \omega^{\frac{1}{2}} V_p / 2Q \kappa^{\frac{1}{2}}$, where V_p is the periodic steady state surface temperature, are shown in fig. 5.

The University of Tasmania.

XIX. *New Contributions to Interferometry. Part III.—The Differential Polarization Phase Change on Reflection at a Thin Silver Film.*

By S. TOLANSKY, Ph.D., D.I.C., The University, Manchester*.

[Received November 1, 1943.]

SUMMARY.

The differential phase change at reflection, for light polarized in and perpendicular to the incident plane, is determined for a silver film. Green mercury light is used, the film being one-tenth of a wave thick. The differential effect is measured interferometrically by a multiple beam Newton's ring apparatus described elsewhere. The effect is measured over an incidence range of 0° to 80° . To the first order the experimental results confirm the predictions of classical electromagnetic theory. Up to an incidence of 60° the experimental curve of the variation with incidence is parallel to the theoretical curve but slightly displaced upwards (by 0.007λ). There exists a marked point of inflection at 63° which is not predicted by theory. The existing theoretical treatment is considered to be incomplete.

The intensity ratios of the two beams, which are separated interferometrically, are visually estimated for different incidences. The beam polarized with the magnetic vector parallel to the incident plane is rapidly absorbed as the angle of incidence increases. The absorption curve exhibits a point of inflection.

The influence of the fringe doubling arising from the differential phase change with the Fabry-Perot interferometer, when used for high resolution spectroscopy, is considered. The doubling leads to a reduction in resolution even before doubling can be directly detected. For interferometers with plate separation less than 1 millimetre, there can result an appreciable loss in resolution.

* Communicated by the Author.

Introduction.

IN Part II. it was shown that the doubling of the Newton's rings formed between silvered surfaces, when the incidence is not normal, is due to the differential phase change which takes place on reflection at a thin metal film, the phase change being different for light polarized in and perpendicular to the plane of incidence. In this paper, the procedure described in Part II. is applied to the quantitative determination of the differential phase change for a specific silver film. The experimental results are compared with the predictions made by applying the classical electromagnetic theory to the particular film thickness employed. The classical theory for the differential phase change at reflection at different angles of incidence has been developed in considerable detail by MacLaurin⁽¹⁾, whose results will be utilized.

It will be shown that there is very excellent general agreement between theory and observation. Although there exist certain quite marked deviations, yet the general trend of the theoretical and observed curves is closely similar. MacLaurin's calculations show that the differential phase change depends upon the thickness of the reflecting film, but, with the particular thickness of silver employed in this investigation, the variation is insensitive to such changes. The differential phase change is only considerably modified by thickness when the latter is a small fraction of a wave-length. Already, when a thickness of one-tenth of a wave has been reached, the theory shows that the relation between the differential phase change and the angle of incidence is quite insensitive to thickness and indeed the curve for this thickness approximates fairly closely to the limit values calculated for massive metal.

This insensitivity to metal thickness in the particular region of the observations made here is fortunate, since an *exact* determination of the thickness presents considerable experimental difficulty. It is possible to obtain a close approximation to the true thickness, and the value so obtained is certainly of sufficient accuracy to enable theory to be subjected to a critical test.

Evaluation of the Film Thickness.

The silver films used were prepared by evaporation *in vacuo* from a specially designed molybdenum filament. A direct determination of the film thickness could perhaps have been made by chemical destruction of the film with estimation of the silver. This was not attempted. The thickness was inferred from an estimate of the reflecting coefficient.

After some experience it is possible to make a good estimate of the reflecting coefficient by inspection of the appearance of the Fabry-Perot fringes formed when the silver is deposited upon high quality optical flats. The fringe width is extremely sensitive to small changes in the reflecting coefficient, when the latter is high. When the reflecting coefficient increases from 0.85 to 0.87, the "coefficient of finesse," using Fabry's concept, increases from 151 to 206, *i. e.* a 36 per cent. increase in fringe sharpening, an amount which is quite easy to detect by simple visual

inspection. An empirical procedure which can be linked with this is to count the multiple images of a lamp filament seen by transmission through a pair of silvered surfaces held, slightly tilted, close to the eye. The sensitivity of this procedure is great. Some 60 images can be seen with a reflecting coefficient of 0.87 (if the transmission is normal). At 0.80 the number visible has fallen to 42, the 60th image having only one five-thousandth of the intensity that it has in the former case. It is impossible to confuse the two, and by this alone the reflecting coefficient, when high, can be estimated to within 0.03 with a fair degree of certainty. It is found in practice that the evaporation procedure is consistent, and some reliance can be placed upon the multiple image test for reflecting coefficient evaluation.

It is considered that the silver films used have a reflecting coefficient of 0.87 for the green mercury line, and it is very unlikely that the error in this value exceeds 0.03.

The reflecting coefficient being known, the film thickness can now be obtained from the relation determined experimentally by Fabry⁽²⁾ between film thickness and reflecting coefficient. This same relationship has been derived from electromagnetic theory by MacLaurin⁽³⁾. A highly satisfactory feature is the considerable degree of agreement between the theory and observation. The curves have the same shape, although the experimental curve is slightly displaced downwards to lower reflecting values. This is as it should be, since any impurities or gas occlusions will lower rather than raise the reflecting coefficient observed. The two curves are sufficiently close to enable the experimental values to be used with confidence. Fortunately Fabry determined his curve with the green mercury line, so that no possible dispersion errors are involved by employing his data.

From both Fabry's data and MacLaurin's theory the reflecting coefficient increases very little when the film thickness of the silver layer is increased beyond one-tenth of a wave (0.1λ). At a thickness of 0.05λ the reflecting coefficient has fallen to 0.79. On the other hand, at thicknesses *greater* than 0.1λ the absorption increases with great rapidity and little light is transmitted. Clearly, then, if transmission is good the thickness cannot exceed 0.1λ by very much. On the other hand, both the multiple reflections and the fringe sharpness show that the reflecting coefficient is certainly much better than 0.79 and cannot be far from 0.87. Thus it can be confidently inferred that the film thickness is about 0.1λ , with an uncertainty not exceeding 20 per cent. either way.

An eminently satisfactory feature in the comparison made later between the experimental determination of the differential phase change and the theoretically predicted values, is that the theory shows that the differential phase change curves for a thickness ranging from 0.05λ to 0.20λ are practically superposed. Even that for 0.02λ , at which the reflecting coefficient is so low as to be rejected without consideration (0.45), differs very little from the curve given for 0.1λ . It is, in fact, only for still thinner films that serious differences are predicted.

It is thus possible to compare, with considerable exactness, the theoretically predicted with the observed data, since the thickness of the film used is known to well within the requisite degree of accuracy.

The Differential Phase Change at Reflection.

It has already been shown in Part II. how the differential phase change between the mutually perpendicularly polarized beams can be obtained from the doublet displacement which appears at non-normal incidence with silvered Newton's rings apparatus. It was found most convenient to project an image of the fringes on to a photographic plate set perpen-

TABLE I.
Differential Phase Change at Reflection.

Angle of incidence (degrees).	Phase change observed.	Phase change theoretical.	Difference.
20	0.017	0.010	0.007
30	0.032	0.025	0.007
35	0.040	0.033	0.007
40	0.055	0.045	0.010
45	0.067	0.060	0.007
50	0.086	0.079	0.007
55	0.110	0.097	0.013
60	0.137	0.120	0.017
61	0.142	0.125	0.017
62	0.153	0.130	0.023
63	0.171	0.136	0.035
64	0.185	0.147	0.038
65	0.190	0.151	0.039
70	0.235	0.191	0.044
75	0.282	0.243	0.041
80	0.345	0.310	0.035

dicular to the optic axis, that is, perpendicular to the direction of the light beam incident on the interference film. The fringes are then in focus over only a few orders, but these suffice. The doubling is best measured along the minor axis of the elliptical fringes which are registered on the plate.

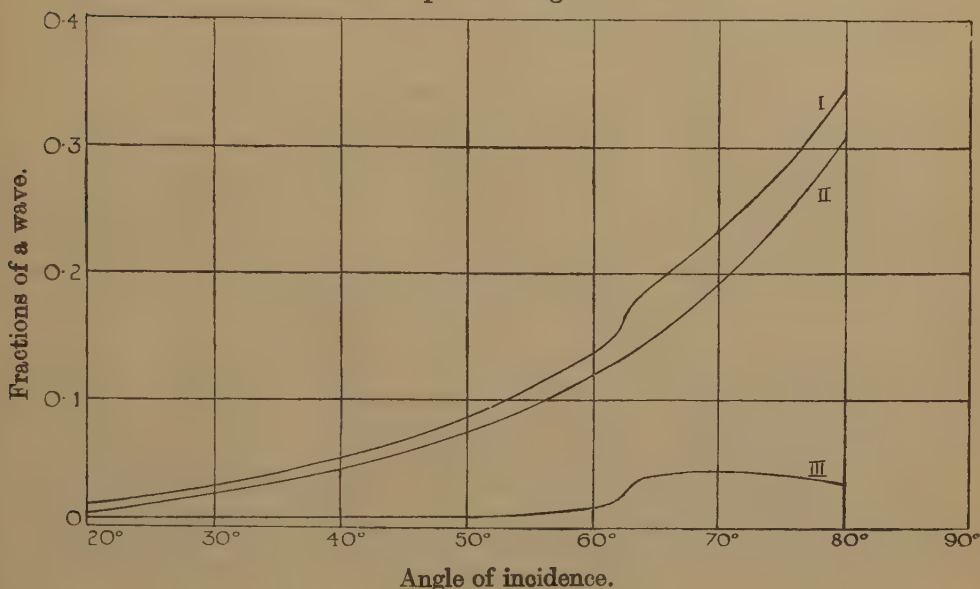
Ferguson ⁽⁴⁾ has shown that the fringes suffer the least distortion from true ellipticity in this direction, but it so happens, too, that the method of reduction employed to evaluate the fraction of an order separation compensates ring distortion effects. The fractional order displacement were determined by the standard hyperfine structure procedure used with off-centre Fabry-Perot fringes, since in both cases the rings obey a parabolic law. The observed changes, as a fraction of a wave, for different angles of incidence, are given in Table I. These are plotted in fig. 1 in the highest curve marked, I.

The precision is high and a very smooth experimental curve is obtained. Measurements were made from 20° onwards, since the doubling could not be resolved for smaller angles of incidence. Observations were carried out at intervals of 5° , except between 60° and 65° , where 1° intervals were used because of the point of inflection found in this neighbourhood.

The theoretical curve, marked II, given by MacLaurin is clearly slightly displaced downwards and *does not show any inflection*. If any error has been made in the estimate of the film thickness it can hardly be shown on this scale. Any error in estimating this is likely to be such as to give a smaller value than the true thickness. If the film is indeed thicker

Fig. 1.

Differential phase change at reflection.



than that estimated, then the discrepancy between theory and observation is still further increased, for the thicker the film the lower the theoretical curve.

This theoretical curve attempts to take into account the transition influences due to the glass supporting surface. *Attention is drawn to the large scale.* The differential phase change is given as a fraction of a wave-length. Up to 60° incidence the experimental and theoretical curves run closely parallel, the theoretical being systematically lower than the experimental by the *very small amount* of 0.007λ . This difference is undoubtedly real, for the experimental precision is high and the uncertainty in the film thickness has a much smaller effect. *Thus to a first order the observations fully confirm the theory.* The difference cannot be considered as of much more significance than a second order effect.

Beyond 60° there is a striking departure from theory, although still on a small scale. The experimental curve shows a marked point of inflection, the position of which was found with accuracy by increasing the observations in this neighbourhood to one per degree. After inflecting, the experimental curve starts roughly parallel with the theoretical, but displaced now by a larger amount, 0.040λ , which is of course still small.

The two curves no longer run parallel but begin to converge slowly as incidence increases to high values.

The difference between the curves I and II is plotted below as curve III. This difference curve *cannot possibly be attributed to any false estimate of film thickness* and certainly represents a real discrepancy between theory and observation. Thus, whilst to the first order, theory is vindicated, *there exists a striking second order difference*, the main characteristic being the point of inflection which occurs in the neighbourhood of 63° .

It is to be noted that 63° is in the region of the quasi-polarizing angle of this silver film.

It is probable that the finer details of the theory are incomplete. A possible source of the discrepancy is the effect of transition at the surfaces. The refractive index does not change abruptly at the front or the back surface of the film. A film 0.1λ thick is perhaps little more than 200 atoms deep, and in no sense can this be considered as uniform throughout its depth. Oxide surface films and glass contamination films doubtless play their part in distorting the surface structures of these thin silver films. The microcrystalline structure is not taken into account in the classical electromagnetic treatment of MacLaurin, and this fact alone may account for the discrepancy. The two curves appear to converge beyond an incidence of 80° and it would have been of interest to pursue this to see whether they meet or cross. This could not be achieved because of the already extremely weak intensity of the variable component at 80° incidence.

The shapes of curves I and II are such that the difference in differential phase change is inconsiderable. On the other hand, the change in the value of the Principal Incidence is quite appreciable. The calculated value for the Principal incidence is $75^\circ 40'$, that observed being $71^\circ 45'$.

The Intensity of the Variable Component.

Visual estimates of the relative peak intensities of the two components have been made. The integrated light in the broader fringes is greater than that in the narrower simply by virtue of the greater width. This factor is not considered, since only the peak intensities are of fundamental interest.

The relative intensity of the variable component falls steadily with increasing incidence. The relative intensities are shown in Table II. In this, the intensity of the main component (that with magnetic

vector perpendicular to the plane of incidence) is taken as the stronger for each angle of incidence.

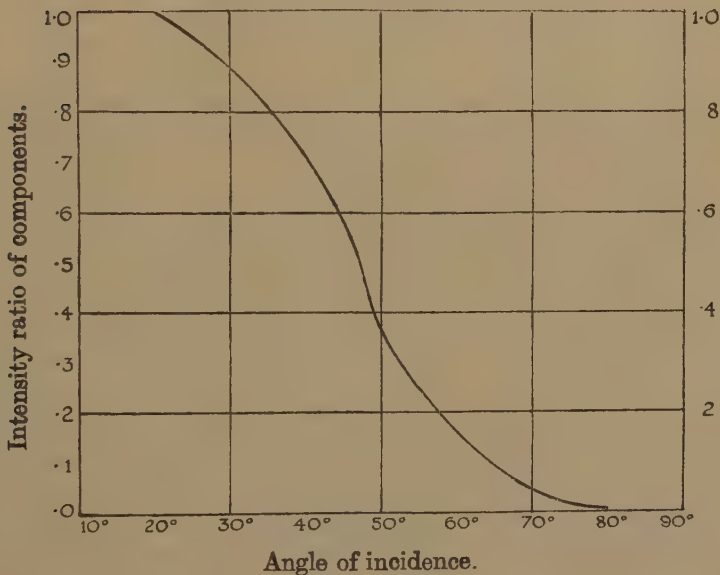
The intensity ratios (smoothed out) are plotted in fig. 2, the resulting curve being of interest. There is a definite point of inflection which is

TABLE II.

Angle of incidence, degrees.	Intensity ratio.	Intensity of variable component as percentage of stronger.
20	1 : 1	100
30	10 : 9	90
35	10 : 8	80
40	10 : 7	70
45	10 : 6	60
50	3 : 1	33
55	4 : 1	25
60	6 : 1	16
65	10 : 1	10
70	25 : 1	4
75	100 : 1	1
80	200 : 1	0.5

Fig. 2.

Absorption of the variable component.



not at the same angle of incidence as the inflection point of the differential phase change curve.

It has already been proved, in Part II. that the variation in the intensity of the two components is to be attributed to the existence of

a differential absorption coefficient for the mutually perpendicularly polarized beams. The absorption coefficient of a metal film is, of course, known to increase with the incidence. The effect shown by fig. 2 is, however, a *differential* effect, since it shows that the *ratio* of the two components changes with incidence too.

Note on the Effect of Differential Phase Change in the Fabry-Perot Interferometer when Used for High Resolution.

It has been shown in Part II. that the differential phase change characteristic of the silvered Newton's rings also appears with Fabry-Perot fringes when viewed at large angles of incidence. The doubling can just be resolved at 20° incidence. Clearly at smaller values the effect is to produce a broadening of the fringes, which means a loss in resolution. It is of interest to calculate the loss in resolution to be expected for the angles of incidence commonly met with in the practical application of the Fabry-Perot interferometer to high resolution spectroscopy, such as for instance the study of hyperfine structure. This is an effect not previously suspected.

In practice the Fabry-Perot is frequently tilted off-centre in order to view rings of perhaps from the 10th to the 16th order. There are reasons for this. The higher order fringes are sharp. With *small* gaps the linear dispersion near the centre is so great that it is advantageous to operate at higher orders, with sharper fringes. Finally, and this is of considerable practical use in reducing the observations, the MacNair and Schrammen standard procedures of reduction can be used with no effective error and the labour of computation is greatly reduced. This is a matter of some value when an extensive spectrum with complex hyperfine structures (for example, iodine or bromine) is under examination. Even when used normal to the incident beam, higher orders are often measured when the gaps are large too.

Simple calculation shows that the phase change effects are only appreciable for small plate separations, but for these, which are just those at which higher orders are invariably examined, the effect definitely exists. Consider a Fabry-Perot interferometer of 1 millimetre plate separation. With the green mercury line the angular radius of the n th ring is $1.34^\circ \sqrt{n}$. This is also the angle of incidence of the light contributing to the n th ring. It is necessary in practice to make observations on high orders. With a typical 1 metre lens (which cannot of course be changed if one is operating within a spectrograph in the parallel beam), the linear diameter of the first ring is nearly five centimetres. With ordinary length spectrograph slits, tilting to bring in higher orders, up to say the 16th, cannot be avoided. Thus a tilt of 5° is required, sometimes more.

The differential phase change effect at 5° incidence clearly affects resolution. At normal incidence there is no differential phase change. The shape of the experimental curve justifies extrapolation down from

20° , which is the first position at which a definite *doubling* can be clearly seen. Broadening sets in, of course, long before doubling into two components can be detected. The extrapolated value of the differential phase change at 5° is such that the components are separated there by about 0.009 of an order (the order separation is double the phase change). The fringes are thus broadened to the extent of 1/100th of an order. This is appreciable when the highest degree of resolution is required. If still higher order rings are observed, or smaller gaps employed, the effect becomes progressively worse.

The remedy is simple. It is necessary only to introduce a Nicol prism or polaroid disc into the beam. Not only does this cut out one component and thus eliminate the broadening, but if in addition the perpendicular component is the one cut out, there is a small *gain* in resolution over that at normal incidence, because of the increased reflecting coefficient producing sharper fringes. There is no effective difference in intensity between the two components at this angle of incidence.

It may be pointed out that the employment of a small gap is often dictated by a widespread structure and, as the resolution is inversely as the gap, any procedure which increases resolution for small gaps is of value. In theory there will always be a net gain with the Fabry-Perot interferometer if the incident light is polarized correctly. There is a loss of light, and for larger gaps it is doubtful whether the small gain, which undoubtedly exists, is worth the light loss. With weak sources one is obliged to use thin films, and a greater improvement in resolution would result from thickening the film to transmit the same light as the polarizing device. The procedure is thus of little value when the light source is weak. If intensity introduces no difficulties, and if the highest resolution is required, the polarization should be adopted, the main object being the elimination of one component rather than the utilization of the rather small increase in reflecting coefficient at small angles of incidence.

With a plate separation of 1 centimetre the doublet separation at the same fringe is only 1/300th of an order, an amount beginning to be too small to be noticeable. Because of the smaller pattern it is usually neither necessary nor desirable to operate far from the 9th order, at which the doublet separation is but 1/450th of an order, and can be disregarded.

It can be concluded that the effect is only of importance to plate separations of the order of 1 *millimetre or less*, the smaller the gap the more serious the effect because the larger the angle of incidence that must be used.

References.

- (1) MacLaurin, Proc. Roy. Soc. A, lxxviii. p. 296 (1906).
- (2) Fabry, 'Les Application des Interferences Lumineuses' (1927).
- (3) MacLaurin, 'Theory of Light' (1908).
- (4) Ferguson, Phil. Mag. xxv. p. 501 (1913).

XX. *Remarks on Electrical Units.*

By D. P. DALZELL*.

[Received November 11, 1943.]

THE paper † in which Prof. Dingle expounds the theory of electric and magnetic magnitudes adds one more to the many schemes for reforming the dynamical system of units, and therefore invites comment from those who are satisfied that this system is philosophically and logically satisfactory. We shall pick out some of Prof. Dingle's remarks which conflict with our view of the matter and summarize the grounds for defence of the existing system.

On p. 590, line 8 up, Prof. Dingle states that the electrostatic and electromagnetic systems of measurement are precisely equivalent, without defining the relation of equivalence between two systems of measurement. On the same page he proceeds to say that in neither system has any clear distinction been made between definitions and experimental relations, and conveys the impression, therefore, that there is no need for definitions of units to take any account of the properties of the things to be measured. This view of the status of units seems to make the activities of the National Physical Laboratory excessively mysterious. Our contention is that the existing scheme of dynamical units takes account of these properties in a skilful and satisfactory manner.

On p. 591, Prof. Dingle quotes the equation

$$f = \frac{q_1 q_2}{kr^2}$$

as though it were used for defining the electrostatic unit of charge. Our view is that no such equation is used. The law of inverse squares is involved in the determination of the dimensions of charge, not in the definition of unit charge.

On p. 591, line 9 up *et seq.*, Prof. Dingle quotes a definition of quantity of electricity which involves the use of an arbitrary standard charge. According to our view this standard charge would be a fundamental unit and would have to be reproduced wherever and whenever an electric charge is measured according to this definition, and a great merit of the existing system of dynamical units is that the preservation of such standard charges is unnecessary. Similar remarks apply to Prof. Dingle's definition of magnetic pole strength, p. 592, line 15.

* Communicated by the Author.

† Phil. Mag. xxxiv. No. 236, p. 588.

On p. 593, Prof. Dingle proposes a definition of unit electromotive force which involves a previous knowledge of the value of the mechanical equivalent of heat. This procedure has in the past been avoided, presumably because the mechanical equivalent was not known to any great accuracy. Furthermore, the occurrence, in the law of electromagnetic induction, of a constant of proportionality which depends on the mechanical equivalent seems to be most undesirable.

On p. 593, Prof. Dingle quotes as a virtue that magnitudes can be obtained by his definitions without drawing on any knowledge acquired by experiment; this appears to us to be a vice, but the definitions themselves are remarkably similar to some of the existing ones which do, in fact, take account of such knowledge. In his remarks on the definition of electric current Prof. Dingle points out that, although his definition does not take account of experimental facts, it is nevertheless fortunately compatible with the fact that a current in a (suitable) circuit is the same at all points along it; he has thus adopted a definition which was designed to have this result in view of the known behaviour of such circuits.

On p. 596, line 7 up *et seq.*, Prof. Dingle states that in a logical development of the subject in which a quantity is known only by the effect which gives its measure, alternative dimensions for the same quantity would not be possible. If this sentence is to be accepted we must agree that a logical development of a subject would be one in which all effects except those adopted for the unique definitions of units would be disregarded, and we would have to forget the transverse and longitudinal Hall effects, magneto-striction, magneto-optical rotation, and the Zeeman effect, in case someone were rash enough to use one of them for the definition of a unit of magnetic intensity, and we would not then be able to talk about magnetic fields in sunspots.

But Prof. Dingle has a more revolutionary scheme; he would invent a new concept for each effect, on the assumption that a concept is something incapable of enlargement by experience. We would prefer to say that we are capable of learning from experience and of enlarging and modifying concepts accordingly, and we are therefore capable of sympathizing with the early investigators who concerned themselves with demonstrating the identity of frictional, voltaic, atmospheric and animal electricity. Prof. Dingle states that the experimental facts of a science, as expressed by the dimensional equations, form a framework which it is customary to clothe with a theory. We would agree with this statement if the phrase "as expressed by the dimensional equations" were omitted, but we would suggest that the construction of theories is not a custom but (among other things) a reasoned policy of reducing the number of concepts. We would also contradict Prof. Dingle by asserting that the electromagnetic theory of light depends on Maxwell's hypothesis concerning displacement current, not upon the scheme of dimensions of units. The latter does, however, confirm the former by providing a numerical value of the velocity of propagation of electro-

magnetic waves, but Maxwell's hypothesis does not affect the dimensions of electrical units and can be omitted from any discussion of them.

In discussing Eddington's statement that the fine structure constant is dimensionless, it is necessary to bear in mind that the meaning assigned to such a statement is a function of the reader's preconceptions. A generous reader would naturally suppose that $ch/2\pi e^2$ is the expression arrived at for the fine structure constant in the system of units now in common use, and that had Eddington wished to anticipate the use of other systems of units he could have done so by including any number of appropriate factors. The value of the fine structure constant $1+2^3+2^7=137$, would then remain unaltered. Eddington clearly is concerned with constructing new theories, not with smashing up old ones. The occurrence of κ or β in the formula for the fine structure constant would have indicated prevision of Prof. Dingle's views on units; the occurrence of c , h and e in that formula is an infallible indication that the results of experiments have been drawn upon. Now we may agree that it is inherently impossible for the dimensionless fine structure constant to be derived from pure reason, but saying this does not deprive Eddington's statements of meaning. We may reasonably suppose that Eddington has constructed a new theory, with new hypotheses, whereby a term in an equation can be inferred by means which do not involve c , h and e , that this term contains a constant occupying the same place as does $ch/2\pi e^2$ in a derivation of the same equation from existing hypotheses. By saying that the value of the fine structure constant may be derived from epistemological principles, Eddington presumably means that new hypotheses seem to him to be much more attractive, fundamental and general than those of existing theories, and one would expect a sympathetic critic to examine Eddington's writings in order to extract a more explicit formulation of these hypotheses and of inferences from them in order to help others to explore the new theory. There is no doubt, in our mind, that too much attention has been paid to Eddington's presentation of his work and too little attention paid to the work itself.

We shall now discuss electrical and magnetic units generally. A fundamental fact in this case is that electricity and magnetism are two separate subjects when stationary field distributions are under consideration. The electrostatic system of units is a dynamical system based on measurements made with stationary electric fields, and the electromagnetic system includes units in a dynamical system based on measurements made with stationary magnetic fields. We shall refer to the dynamical system of units based on measurements made with stationary magnetic fields as magnetostatic units. Thus, in our nomenclature, electrostatic units include electric charge, electric intensity, electric polarization and electric potential. Magnetostatic units include magnetic pole, magnetic intensity, magnetic induction, unvarying electric current. There is no way of measuring electric current and magnetic intensity in electrostatic units and no way of measuring electric charge and electric intensity in magnetostatic units. In order to study the conduction of

unvarying currents a provisional comparison unit based on the electromotive force of a voltaic cell can be adopted. Prof. Dingle's suggestion in this connection yields a thermal unit of electromotive force. When varying fields are studied electricity and magnetism unite, and each system of units is extensible over the field of the other. We shall refer to the extended electrostatic system as the electrokinetic system and to the extended magnetostatic system as the magnetokinetic system. The new results obtained with varying fields are that electric current is proportional to the rate of variation of charge and that electromotive force is proportional to the rate of variation of flux of induction. In both systems the measure of current is taken to be equal to that of rate of variation of charge, and electromotive force is taken as equal in magnitude to the rate of variation of flux of induction. Thus the duality of units seems to be perfectly natural, not a whim, but a consequence of the dichotomy associated with stationary fields, and is an instructive exemplification of the fact that dimensions are attributed to physical magnitudes by a system of measurement. The situation seems to have been obscured by a not very satisfactory nomenclature.

A point on which all reformers seem to concentrate is the measurement of charge in the electrostatic system. The existing definition is that unit charge is such that if two bodies, each having unit charge, are sufficiently small and are separated in a vacuum by unit distance then unit mechanical force is exerted upon each by the other. This is a dynamical unit because the unit charge is defined in terms of a dynamical measurement. The definition naturally assumes that the comparison of charges has already been established, and that therefore the relation of "having the same charge" is known.

The dimensions of unit charge follow not merely from the definition of the unit but jointly from the law of inverse square and from the dimensions of dynamical units; for the law of inverse squares tells us how the magnitude of the unit charge varies when the unit of length is varied, and therefore also the distance between the two unit charges in the definition. We recall that the dimensions of a derived unit are a statement of the dependence of the magnitude of the unit upon the magnitudes of the fundamental units, and from this fact it should be obvious that the dimensions of a physical magnitude depend not only upon the system of measurement but also upon experimental information. We shall call a system such as that of two unit charges separated by unit distance a unitary model. Thus, a change of fundamental units changes the scale of the unitary models. This fact has a bearing on the relation between the method of similarity and that of dimensional homogeneity. The method of similarity compares two systems of different scales while the method of dimensional homogeneity keeps the system under consideration the same but amounts to applying the principle of similarity to the unitary models. In Prof. Dingle's definition of charge one would naturally suppose that having assumed that an arbitrary standard charge can be chosen the simplest procedure would be to adopt it as a unit of comparison.

XXI. *The Ultrasonic Diffraction Grating.*

By R. A. HOUSTOUN, M.A., D.Sc., University of Glasgow *.

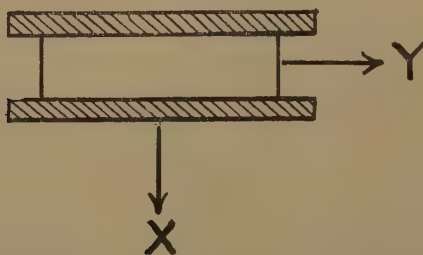
[Received January 5, 1944.]

Summary.

This paper describes the quartz ultrasonic diffraction grating and some purposes for which it can be used. It is based on the experimental work which has been done with it at Glasgow.

WHEN a quartz plate is placed in a periodic electric field, stationary ultrasonic waves are formed in it. It then becomes a piezo-electric oscillator. If the period of the waves coincides with one of the natural

Fig. 1.



periods of the quartz, these waves may be relatively large. They consist of rarefactions and condensations; a rarefaction diminishes the refractive index and a condensation increases it, so the different layers of the quartz have different refractive indices, the refractive index varying harmonically in the direction of propagation of the wave. Consequently, if a parallel beam of light traverses the quartz in a direction at right angles to the direction of the ultrasonic waves, a periodic change of phase is produced, which causes a diffraction spectrum.

The matter will be made clear by the consideration of an actual case (fig. 1). Here there is a rectangular quartz in position between condenser plates; the lengths of the X, Y and Z sides are respectively 3.52, 12.975 and 0.955 mm., OZ is parallel to the optic axis and OX is parallel to an electric axis of the quartz. When the electric field is applied, there is a stress both in the X and in the Y direction. If the frequency coincides with one of the thickness frequencies, we have thickness vibrations; if it coincides with one of the length frequencies, we have length vibrations.

* Communicated by the Author,

The broken lines in figs. 2 and 3 show the positions of the nodes for the third thickness and third length harmonics respectively. As the two plates of the condenser have charges of opposite sign, the upper surface of the quartz contracts when the lower expands and vice versa. Consequently only the odd thickness harmonics are present. Both thickness and length harmonics give diffraction bands; as the frequency of the electric field is gradually increased, whenever it comes into coincidence with a natural frequency, the corresponding bands flash into view, and, since for the crystal in question the length is nearly four times the thickness and the velocities in the X and Y directions do not differ much, the length harmonics occur about four times as frequently as the thickness harmonics. The length bands are not as bright as the thickness bands and it is better to use only the latter.

Fig. 2.



Fig. 3.



The following quartzes have given bands with most of the following oscillators :—

	<i>Quartzes.</i>			
A	1.02	× 25	× 22	mm ³ .
B	2.535	× 15	× 9	„
C	1.98	× 15	× 10	„
D	3.52	× 21.95	× 12.975	„
E	5.88	× 21.95	× 12.975	„
F	12.00	× 22.01	× 12.98	„
G	3.435	× 22.01	× 12.98	„
H	2.05	× 12.98	× 22.01	„
I	0.99	× 12.98	× 22.01	„
J	3.52	× 9.55	× 12.975	„
K	3.54	× 60	× 13.01	„

In every case the dimension given first is the thickness. It is parallel to a polar axis of the quartz and is the direction in which the electric field is applied. The length is given second; it is parallel to the optic axis and is the direction in which the light traverses the quartz. An AT cut-crystal, 1.06 mm. thick, would not give bands.

Oscillators.

(1) A push-pull oscillator using Osram DET 12 valves. The quartz was placed between the condenser plates and tuning was done by varying the induction. The range of wave-lengths was about 276–373 cm.

(2) A push-pull oscillator using Standard 4316 A valves. In this case the quartz was placed in the condenser of a closely coupled secondary and the range of wave-lengths was about the same as for (1).

(3) A push-pull oscillator using Mazda P 220 valves. Tuning was done by varying the capacity and the quartz was placed between the condenser plates of a closely coupled secondary. The range of wave-lengths was about 470–762 cm.

(4) A push-pull oscillator using Standard 4316 A valves. The quartz was placed between the plates of the condenser and tuning was done by varying the induction. This went down to a wave-length of 184 cm.

(5) A Hartley oscillator using a Standard 4316 A. The quartz was placed between the plates of the condenser. This had an electrical wave-length of 140 cm. and the bands it gave were feeble.

When a push-pull oscillator is used, it seems necessary for the production of good bands that the grid current should be high; they are obtained with a low anode voltage and nothing is gained by increasing the voltage.

A simple and satisfactory theory of the production of the bands has been given in a series of papers by Raman and Nath ⁽¹⁾. According to this theory, if λ_1 is the wave-length of the ultrasonic wave in the quartz and θ the deviation of the first order spectrum

$$\sin \theta = \frac{\lambda}{\lambda_1}.$$

For the second order spectrum

$$\sin \theta = \frac{2\lambda}{\lambda_1},$$

and so on, just as for the ordinary diffraction grating. With oscillator (1) the crystal represented in fig. 1 works best on its 137th harmonic; it is then equivalent to a grating of 137 rulings. In the case of crystal F I have worked regularly on the 475th harmonic; this is equal to a grating of 475 rulings. Considered as gratings these quartzes are excellent; owing to the small number of rulings the resolving power is not high, but under favourable conditions F has been able to separate the yellow Hg lines and, under favourable conditions, the bands are bright enough to project on a screen and show to a class.

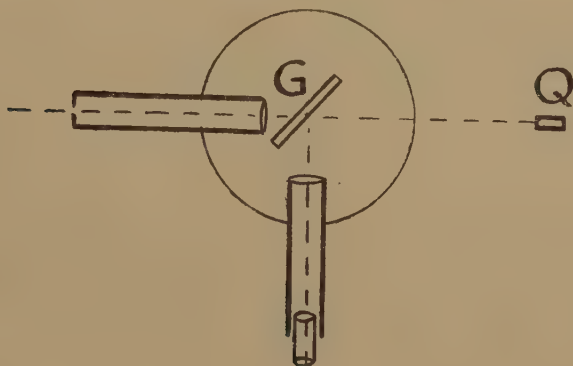
The easiest way of observing the bands is shown in fig. 4. Light from the collimator passes through a glass plate G, enters the quartz Q and is reflected by its second surface; it then returns on its path and is reflected by the plate G, so as to enter the telescope and the eye of the observer. Less than one per cent of the initial energy survives the double reflection, but with a Hg or Na arc as source there is light enough and to spare. But, if the bands are faint, collimator and telescope must be set up in line, on separate stands, and the oscillator placed between them. With the arrangement depicted in fig. 4 the three-metre oscillators (1) and (2) give first and second order spectra on each side of the direct image, the deviation of the first order being about 32', and the six-metre oscillator (3) gives three and sometimes four spectra on each side, the deviation of the first order being then half as great.

At these high harmonics the quartz does not react back on the oscillator appreciably. For example, on one occasion with oscillator (1) and crystal J at the 137th harmonic the anode current was 45 mA. and the grid current 16 mA. On passing through resonance there was no appreciable change of the anode current and $\frac{1}{10}$ mA. could have been detected; on passing through resonance there was a slight variation of the grid current, but not more than $\frac{1}{4}$ mA.

The quartz ultrasonic grating may be used (a) as a grating, (b) as a means of measuring the velocity of light, (c) as an electric wave-meter, and (d) as a means of measuring the persistence of luminosity in an atom.

As a Grating.—The low resolving power limits its use in this direction. The rulings must be very regular, as they are spaced out by the laws of wave-motion and not ruled by an engine, but there is apparently no

Fig. 4.



purpose for which this would be an advantage. If the electrical wavelength were diminished to 3 cm., the resolving power would be comparable with that of the best ruled gratings, but the ultrasonic wavelength in quartz would then be about 5×10^{-5} cm., entailing a quite impossible degree of parallelism in the incident beam.

But the theory of the grating enables us to draw conclusions regarding displacements of the particles in the ultrasonic wave-front from the brightness of the bands. If the refractive index for a given time is taken as

$$\mu_0 - \mu \sin \frac{2\pi x}{\lambda_1},$$

the ratio of the amplitude of the first order beam to the amplitude of the direct beam is given by $J_0(2\pi\mu l/\lambda)$, which becomes $2\pi\mu l/\lambda$ when the ratio is small; here l is the length of the path in the crystal and λ the wave-length of light. Since μ varies harmonically, the mean ratio of the intensities, when it is small, is

$$\frac{1}{2} \left(\frac{2\pi\mu l}{\lambda} \right)^2.$$

For crystal J, when this is one-tenth, μ works out as $\cdot 0745 \lambda$. Thus, for Na light the root mean square value of μ/μ_0 is $2\cdot 84 \times 10^{-6}$. Assume that this is equal to the strain du/dx and, since we are dealing with mean values, put the latter equal to $2\pi u/\lambda_1$. For crystal J and the 137th harmonic $\lambda_1 = \cdot 00514$ cm. Hence the root mean square value of the displacement is $\cdot 00514 \times 2\cdot 84 \times 10^{-6}/2\pi = 2\cdot 3 \times 10^{-9}$ cm.

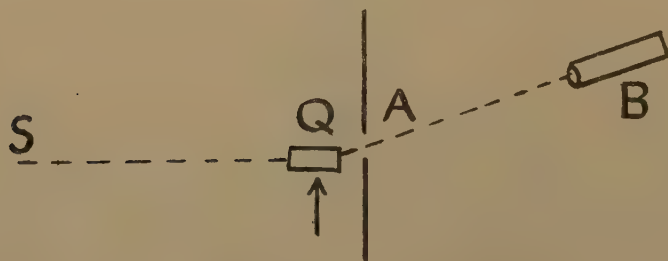
Under static conditions the strain should be given by $d_{11}E_x$ where d_{11} , the piezo-electric constant, is $2\cdot 5 \times 10^{-12}$ metre per volt and E_x is the field strength. If the potential difference between the plates is 1 volt, since they are 3.52 mm. apart, the strain would be

$$2\cdot 5 \times 10^{-12}/3\cdot 52 \times 10^{-3} = 7\cdot 1 \times 10^{-10},$$

one four thousandth of the value obtained above. Of course the potential difference and ratio of the intensities are greater than the values taken and there is the increase due to resonance. But the result, such as it is, shows that there is an interesting field for calculation here.

Measurement of the Velocity of Light.—In fig. 5 let Q be the quartz, SQ the direction of the incident light, and let the arrow denote the

Fig. 5.



direction of the electric field. QA gives the direction of one of the first diffracted beams, the other beams being stopped by the screen at A. B is a lens-mirror combination which reflects the beam back on its path. If, when it returns to Q, the grating is in action, the first order spectrum is diffracted back along QS; if, on the contrary, the quartz atoms are in their mean positions, there is no diffraction. Consequently when the lens-mirror combination is drawn back, the return beam undergoes a periodic change of intensity and the arrangement affords a means of measuring the velocity of light. It replaces Fizeau's toothed wheel by an intermittent grating which interrupts the light 200 times as rapidly as the toothed wheel does, thus greatly increasing the power of the method.

I have recently measured the velocity of light in water by this method, using a length of 4 metres of water which was traversed twice. Three different colours were used and the accuracy obtained was one part in seven hundred. An account of this work is being published elsewhere ⁽²⁾.

The grating interrupts the beam twice in each electrical period. So, if the electrical wave-length is three metres, the positions of the reflector which give minima are 37.5, 112.5, 187.5, 262.5 cm. etc. from the quartz, and if the fundamental frequency of the quartz and the number of the harmonic are known, a fair determination of the velocity of light in air can be made in a small room.

As a Wave-meter.—Bergmann has described an instrument which uses the deviation of the first order diffraction band as a means of measuring electric wave-length. The length overtones of the quartz plate deviate from the harmonic law, but Bergmann⁽³⁾ has shown by careful measurements on six quartzes that, when the thickness is small compared with the other dimensions, the harmonic law is obeyed very accurately by the thickness vibrations. The following table gives some information about these measurements :

No.	Dimensions of plate in mm.			Fundamental frequency		Range of harmonics observed.
	l_x	l_y	l_z	Measured.	Calculated.	
1	9.99	19.97	19.95	281.69×10^3	288.18×10^3	3-191
2	5.78	55.0	55.0	497.5	498.97	2- 53
3	2.33	25.0	25.0	1233.5	1238.7	1- 25
4	1.14	15.0	15.0	2523.3	2531.2	1- 13
5	6.5	15.0	15.0	7- 71
6	2.1	15.0	15.0	5- 23

The wave-lengths were measured with a resonance wave-meter and in the case of the ultra short waves with a Lecher bridge, and the deviation of the diffraction image was measured with a microscope. If the first three harmonics of (1) are excepted, both wave-lengths and deviations obey the law to less than one-tenth per cent. Bergmann suggests that the departure from the law in these three cases is due to the greater intensity of the oscillations setting up lateral vibrations. The fourth column gives the measured frequency of the fundamental in cycles per second and the fifth the value given by the harmonic law. There is agreement except in the case of (1). The velocity of the ultrasonic wave given by the thickness and the calculated fundamental frequency is 5762, 5772, 5772, 5771 m. per sec. for quartzes (1) to (4) respectively, the mean being 5769 m. per sec. The deviation, of course, varies with the electric wave-length, but when reduced to a wave-length of 3 m. it is for the first three quartzes 31.34', 34.10' and 34.06' respectively. These values are for white light. Bergmann does not give the absolute value of the deviation for the other quartzes.

It follows as a consequence of the overtones obeying the harmonic law, that the velocity of the ultrasonic wave is constant over the range

of frequency investigated. As the electric wave-length is decreased, there must come a point where this velocity changes. But according to Bergmann this is below 5 m.

For crystal B, using the arrangement of fig. 5 to determine the wave-length and observing on the fourth minimum, I obtained the following results :—

No. of harmonic.	Fundamental wave-length.
103	2.717×10^4 cm.
95	2.720 „
87	2.726 „
77	2.720 „

The second column gives the fundamental wave-length calculated from the harmonic in question. The mean value is 2.721×10^4 cm. Hence the velocity of the wave works out at

$$\frac{3 \times 10^{10} \times 2 \times 2.535}{2.721 \times 10^4 \times 13^3} = 5590 \text{ m. per sec.,}$$

and the result can be relied on to one-tenth per cent. For another range, with oscillator (3), the same crystal gave the results in the first three columns of the following table :—

No. of harmonic.	Wave-length.	Deviation.	No. of harmonic.	Wave-length.	Deviation.
	m.			m.	
41	6.52	16.5'	29	7.12	14.5'
43	6.24	17.3'	33	6.26	18.0'
45	5.90	17.7'	35	5.80	19.5'
47	5.76	19.2'	37	5.44	19.5'
49	5.40	20.0'	39	5.20	21.0'
51	5.28	21.0'	41	5.00	21.5'
53	5.06	21.5'			
55	4.92	21.5'			

In this case the wave-lengths were determined by a Lecher bridge and the deviation by a spectrometer. The value of the velocity obtained from the mean value of the fundamental wave-length is 5672 m. per sec. The difference from the former value is due to the fact that in this case the wave-lengths were determined with the secondary of the oscillator uncoupled. If the deviations are reduced to 3 m., the result is 35.9'; this is for Na light. If we correct for the wave-lengths being too small, the result is 36.4'.

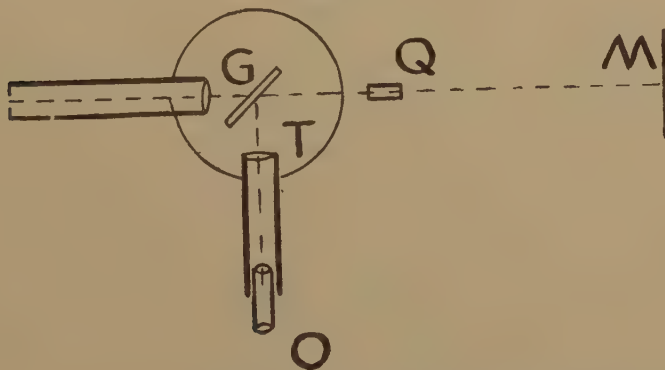
The second three columns of the table give results obtained for crystal C in the same way. The velocity and the reduced deviation worked out from these results are 5819 m. per sec. and 36.2' respectively.

With crystal J and oscillator (1) the electric wave-length was found by the method of fig. 5 to be 283.28 cm. for the 137th harmonic and the deviation measured by a microscope was found to be 31.74' for Hg green, the results being accurate to one-tenth per cent. This makes the velocity 5442 m. per sec. and the deviation reduced to 3 metres 29.98' for Hg green. For Na yellow it would be 32.35'.

The deviations of the harmonics from the 103rd to the 139th were measured for crystal D and were found to obey the harmonic law within the error of observation, which was about one half per cent.

Taking all the observations together, Bergmann's and my own, there is no doubt that the harmonic law is true both for the wave-length and the deviation, within the limits he states. But there is also no doubt that the velocity of the ultrasonic wave, and consequently the grating constant, vary from quartz to quartz. Whether this is due to the circumstances under which the crystal grows or to inhomogeneity in its

Fig. 6.



substance or whether it is conditioned by the other dimensions of the quartz, it is impossible to say.

The mathematical investigation of the oscillations is difficult, but if the arrangement of fig. 6 is used, it is possible to see what is going on. Here the light passes through the glass plate G, the quartz Q, and is reflected back on its path by the mirror M. It then traverses the quartz a second time, is reflected by G and enters the eye of the observer at O. Only the direct light gets through the quartz the second time, the diffracted images formed by the first passage being thrown to the side.

Let us suppose that the quartz is in its 137th harmonic and that the oscillations are represented by $\cos 2\pi\sigma t \cos m\pi x/l$, where $m=137$. Then there are 137 nodes parallel to the YZ side; these are shown in fig. 7. The refractive index varies harmonically in the x direction. Only at the loops do the rays go straight through; at the nodes there is a prismatic deviation. The telescope is focused on the quartz by the long path, i. e., the object distance is $QM+MQ+QC+GT$. If the resolving power were sufficient, the observer would see these nodes; the arrangement

here is an application of Toepler's striation method and on the return journey the quartz acts merely as an aperture. But they are too close together and remain invisible.

What he does see, if resonance is approached from the side of higher frequency, is, in succession, figs. 8 *a*, *b*, *c*, *d*, *e*, that is, 5, 4, 3, 2, 1 nodes and then uniform illumination, the intensity increasing from fig. 8 *a* to fig. 8 *e*; when the frequency is diminished a little more, the quartz goes out of resonance. The case *e* is very rarely seen, as the tuning is scarcely fine enough to bring it out; there is a jump from two nodes to none.

Fig. 7.

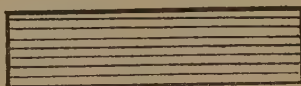
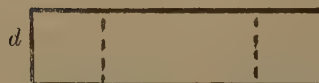
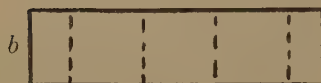


Fig. 8.



Clearly the oscillation is represented by

$$\cos 2\pi\sigma t \cos \frac{m\pi x}{l} \cos \frac{n\pi y}{b},$$

where b is the breadth, and n diminishes from 5 to 0. If we take a for the velocity of the wave and substitute in the equation of wave motion, we obtain

$$\sigma^2 = \frac{a^2}{4} \left(\frac{m^2}{l^2} + \frac{n^2}{b^2} \right) = \frac{m^2 a^2}{4l^2} \left(1 + \frac{n^2 l^2}{m^2 b^2} \right)$$

or

$$\sigma = \frac{ma}{2l} \left(1 + \frac{n^2 l^2}{2m^2 b^2} \right),$$

since the second term in the bracket is small. If the crystal is J,

$$\frac{l^2}{2m^2 b^2} = \frac{1}{2} \left(\frac{3.52}{137 \times 12.975} \right)^2 = 1.96 \times 10^{-6}.$$

So, if m_0 is the frequency of the 137th harmonic, the quartz goes into oscillation at a frequency slightly greater than

$$m_0(1 + 25 \times 1.96 \times 10^{-6}),$$

the intensity gradually increases until the frequency is m_0 and immediately below this the crystal goes out of oscillation altogether.

It is probable also that the displacement varies with z , its complete representation being

$$\cos 2\pi\sigma t \cos \frac{m\pi x}{l} \cos \frac{n\pi y}{b} \cos \frac{p\pi z}{d},$$

where d is the length in the z direction. The reflector can be drawn back to give minima when there are diagonal oscillations present, *i. e.*, when n and p are not zero; but the minima are then not quite so sharp.

One of the crystals together with the condenser plates on each side of it was put in an airtight enclosure and the pressure reduced to 1 cm. of Hg. This made no change in the frequency of resonance, in its intensity, nor in the behaviour of the vertical nodes.

Persistence of Luminosity.—The rate at which an oscillating electron radiates energy is given by

$$I = I_0 e^{-2\alpha t},$$

where α is obtained from the theory of the Hertzian oscillator. As the wave is damped, it can be represented by a Fourier integral and gives a spectral line of finite breadth, the breadth being given in terms of α . The theory has been verified by measurements on the breadth of absorption lines.

W. Wien⁽⁴⁾ has attacked the problem in another manner. Canal rays were passed through a narrow slit into a very high vacuum and the decline of energy along the beam measured by photographic photometry. The velocity of the rays was obtained by the Doppler effect. Hence α could be determined, and for the three hydrogen lines H_α , H_β and H_γ it was found to have the value 5.4×10^7 sec.⁻¹. Thus the intensity would diminish to one-twentieth of its value in 2.8×10^{-8} sec.

When a glass bulb containing gas at low pressure is placed close to the condenser plates of an oscillator, it lights up. This is the electrodeless discharge; no electrodes are necessary, for the light travels into the bulb as electromagnetic radiation. According to a theory given by J. J. Thomson⁽⁵⁾, it is the impact of free electrons speeded up by the field which makes the atoms luminous.

Let us suppose, now, that the arrangement of fig. 5 is used and that a glass bulb containing neon, helium or some other gas at a suitable

pressure, is placed immediately above the quartz, also that the reflector is drawn back and the bulb viewed through the quartz. If the discharge is intermittent, *i. e.*, if there is a flash in the bulb twice in each electrical period, minima will be obtained as the reflector is drawn back. But if the flash persists longer than three or four half periods, the luminosity should be appreciably constant and there should be no minima. I tested several bulbs in this way, using an electrical period of 10^{-8} sec. and obtained no effect. This is what is to be expected from Wien's result.

But if the frequency of the oscillator were gradually diminished, the minima should be obtained eventually, and, from the way in which the depth of the minimum varies with the frequency, it should be possible to calculate α . The experiment would be difficult, but it would supply an interesting check on Wien's results.

I am indebted to my colleague Mr. A. F. Howatson for some assistance with the measurements recorded in this paper.

References.

- (1) Raman, C. V., and Nath, N., *Proc. Ind. Acad. Sci.* ii. pp. 406, 413 (1935) : iii. pp. 75, 110, 459 (1936).
- (2) Houstoun, R. A., *Proc. Roy. Soc. Edin.* lxii. p. 58 (1944).
- (3) Bergmann, L., *Ann. d. Phys.* (5) xxi. p. 553 (1935) ; *Z. Hochfrequenztechn.* xlviii. p. 87 (1936).
- (4) Wien, W., *Ann. d. Phys.* lx. p. 597 (1919) ; lx. p. 229 (1920) ; lxxiii. p. 483 (1924).
- (5) Thomson, J. J., *Phil. Mag.* x. p. 280 (1930).

XXII. *A Non-recording Microphotometer with Variable Contrast Sensitivity.*

By T. B. RYMER, Ph.D., and C. C. BUTLER, B.Sc., The University, Reading †.

[Received December 13, 1943.]

ABSTRACT.

A Hilger measuring microscope has been converted into a non-recording microphotometer capable of measuring the positions of broad lines to 1μ and measuring line intensities. For the measurement of line intensities the electrical circuits are so adjusted that the response of the instrument is linear. When the position of a broad line is to be accurately measured, the constants of the circuit can be quickly adjusted so that the peak of the line is much exaggerated, thus greatly reducing its apparent width.

† Communicated by Professor J. A. Crowther, M.A., Sc.D.

OF the many microphotometers which have been described, by far the greater number have been of the *recording* type, in which a graph is automatically plotted of the density of a point on a photographic plate against the co-ordinates of this point. One might therefore be tempted to conclude that a non-recording microphotometer is necessarily more inconvenient than one of this type. This, however, is not the case, and for certain purposes the recording instrument may be less suitable than its simpler counterpart. The information which is commonly required from a photograph is a table of the distances of the density maxima from some reference mark, together with the values of these densities. The record produced by a recording microphotometer does not usually give this information directly: it has to be measured up. It is therefore apparent that if a photograph consists only of a few lines, less time will be required to measure the positions of these with a non-recording instrument than to obtain a record with a recording instrument, which will then have to be measured.

The *density* of a region on a photographic plate is defined by

$$D = \log_{10}(I_0/I),$$

where I_0 is the intensity of light falling on the plate and I is the amount of light transmitted. The value found for D depends on whether it is measured using a parallel beam of light, a convergent beam or diffuse illumination (Callier † effect). The effect of changing the method of illumination is to multiply all densities on the plate by the same factor. The finer the grain of the plate, the smaller is the variation of D with the conditions of illumination. In discussing the behaviour of a microphotometer, it is convenient to define an analogous quantity D^* by the equation

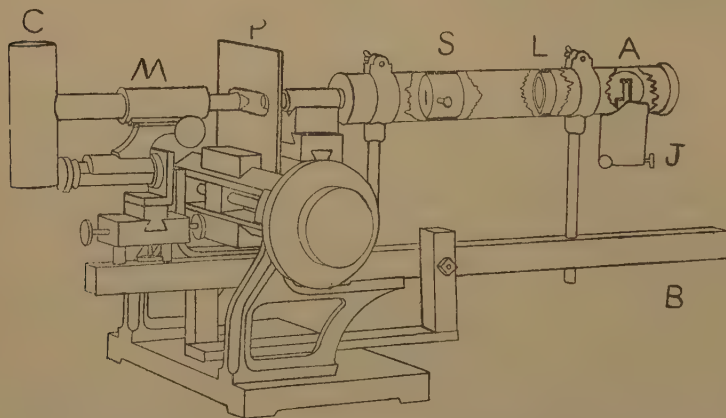
$$D^* = \log_{10}(\theta_0/\theta)$$

where θ_0 is the galvanometer reading when the plate is removed and θ is the reading when the plate is present (the galvanometer is assumed to read zero when no light falls on the photo-cell). In general, D^* differs from D , partly because of the Callier effect (D is understood to be measured with parallel light, whereas a microphotometer uses highly convergent light), but mainly because the galvanometer reading may not vary linearly with the amount of light falling on the photo-cell. In what follows we will ignore the Callier effect, which is negligible if fine-grain plates are used, and in any case may be thought of as merely modifying slightly the characteristic curve of the photographic plate. Then, if the galvanometer reading varies linearly with the amount of light falling on the photo-cell, $D^* = D$; if the deflection varies as the n th power of the light intensity, $D^* = nD$; in general, a graph of D^* against D will be a smooth curve. When a valve amplifier is interposed between the photo-cell and galvanometer, the form of the $D^* - D$ curve can be to some extent controlled by suitable design of the circuits, and we are led to consider the best form of this curve.

† A. Callier, *Zeit. wiss. Photogr.* vii. p. 257 (1909).

There are two main uses of a microphotometer. One is to determine the *intensity* of a line or its contour, and the other is to obtain a more precise measurement of the *position* of a line than is possible by visual observation with a measuring microscope. For the first type of measurement it is advantageous if the range of densities on the plate corresponds to galvanometer deflections between full-scale and, say, one-tenth of this. The percentage error in reading the galvanometer, and, therefore, in determining the line intensity, rises rather rapidly for deflections below about one-tenth of full-scale deflection. Hence if $D^*=D$, the instrument is very suitable for density measurements provided the maximum density is less than 1. If $D^*=nD$, where $n < 1$, intensity measurements can still be made, but the change in D^* , and, therefore, in the galvanometer reading, in traversing a line is smaller and the instrument is not so suitable for weak lines. For measuring the *position* of a line, it is helpful to

Fig. 1.



Mechanical construction of microphotometer.

have an S-shaped curve connecting D^* and D , so that a small change of D about its value at the centre of the line corresponds to a change of D^* from nearly zero to about unity. On traversing a line, the galvanometer will remain at nearly full-scale almost up to the centre of the line, and will then give a sharp dip in passing through the actual centre, which may, therefore, be accurately located. It is one of the virtues of the instrument to be described that the electrical circuits can be instantly adjusted so that the curve connecting D^* and D is either of the form $D^* = \frac{1}{4}D$, permitting intensity measurements to be made on the stronger lines, or an S-shaped curve, with its greatest slope at any value of D between 0.05 and 0.9. By a slight change of the circuit, it can be arranged that $D^* \approx D$, which is most suitable for intensity measurements. Some of the possible D^*-D curves are shown in fig. 3.

The mechanical construction of the instrument is shown in fig. 1, and follows the usual principles. A Hilger measuring microscope forms the

basis of the instrument. The drum of this was graduated to read to 10μ (0.01 mm.). We found the errors of the screw nowhere exceeded 1μ , and, therefore, fitted a vernier enabling readings to be taken to this limit of accuracy. The microscope was removed from the instrument, and the vertical plate-holder P was mounted in its place. This was made of $\frac{1}{8}$ -in. brass sheet, and was machined so that the plate could move accurately in its own plane. This adjustment was checked by focusing a microscope fitted with a $\frac{3}{8}$ -in. objective on to the plate and observing that the plate remained in focus as the plate carriage was traversed from end to end.

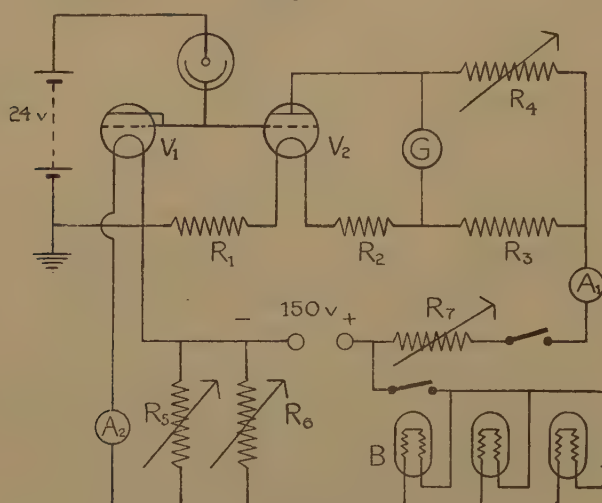
Light from a 24-watt automobile lamp A is focused by a lens L on to an adjustable slit S (obtained from an old spectrometer). The image of the lamp filament can be accurately aligned along the slit by means of the adjusting screws J. The slit is imaged on to the photographic plate by means of a $\frac{3}{8}$ -in. microscope objective of N.A. 0.28. For the usual adjustment of the slit its image has the approximate dimensions $900\mu \times 25\mu$. The light transmitted through the plate is collected by a microscope M fitted with a similar $\frac{3}{8}$ -in. objective. The photo-cell housing C fits into this microscope in place of the usual eyepiece, and is so designed that if the microscope is focused on to the plate using the ordinary eyepiece, then the plate will be imaged on to the photo-cell cathode when C is inserted in place of the eyepiece. The whole optical train is carried on a massive iron bar bolted to the travelling microscope framework, and provision is made for lateral and vertical adjustment of the various components during the initial setting up of the instrument.

We have sometimes found it necessary to make measurements of angle on the plate. To do this, we make use of a large ball-bearing having a central aperture of $2\frac{1}{2}$ in. diameter. The outer housing of the bearing is clamped to P and the inner one carries a circular plate with a scale of degrees round its circumference and having a large circular aperture. The plate is carried on this circular disc. At the moment a paper scale of half-degrees only is used, but commercial ball-bearings are made to such high standards of precision that it would be permissible to engrave a scale fitted with verniers reading to $1'$ should this ever be necessary.

The electrical circuit is shown in fig. 2. The only photo-cell available was an Osram type CMG8, and this has proved perfectly satisfactory. With the optical system used, the maximum current which the photo-cell produces is about 10^{-7} A., and, therefore, an extremely sensitive galvanometer or an electrometer is needed for accurate measurements. We chose the second alternative. The photo-cell current passes through the valve V_1 , which is arranged as a diode. The resulting potential difference across V_1 is applied in the grid circuit of the electrometer valve V_2 and causes a change in the anode current of this valve which is recorded by the Unipivot microammeter G. This has a range of $120-0-120\mu\text{A.}$, and can be shunted as required. The normal anode current of this valve can be backed off by adjustment of R_4 . V_1 is an AT40 valve. This is

a valve with a pure tungsten filament, and is now obsolete; specimens can, however, still be obtained. We have tried a number of valves with oxide-coated filaments, and these were all unsuitable. On the other hand, an "R" valve (another obsolete valve with a pure tungsten filament) is a suitable alternative, though requiring a different filament current. On one or two occasions the AT40 valve has failed to work properly, due to the filament becoming contaminated, but a cure was effected by flashing the filament for a few seconds at a high temperature, thus evaporating off the contamination. This treatment, even if it does not completely restore the original characteristics of the valve, certainly restores its efficiency in the microphotometer.

Fig. 2.



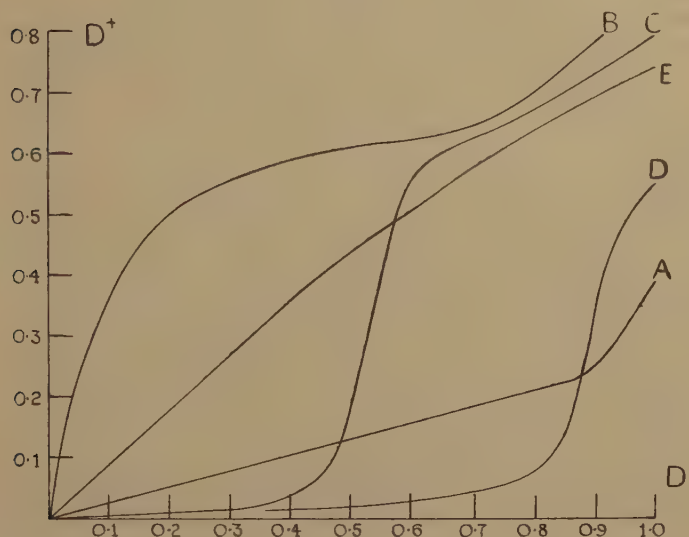
Electrical circuits.

V_1 . AT40.	R_5 . 16 Ω max.
V_2 . Osram type T electrometer triode.	R_6 . 500 Ω max.
B . Osram type 303.	A_1 . 0/0.1 A.
R_1 . 20 Ω .	A_2 . 0/1 A.
R_2 . 60 Ω .	G . Unipivot 120-0-120 μ A.
R_3 . 80 Ω .	Resistance 50 Ω .
R_4 . 20,000 Ω max.	

The filament current of V_1 has to be well stabilized and very closely controlled. As our laboratory has all rooms wired from a 150-v. accumulator battery, we used this to supply the filament, and further stabilized the current by means of barretters B ; coarse and fine control are provided by R_5 and R_6 . The filament current of V_2 is much less critical, and we obtain it from the 150-v. main through the resistance R_7 . About 20 minutes must be allowed for the components to reach thermal equilibrium, when drift becomes negligible and short-period variations of anode current are less than 1 per cent. of full-scale,

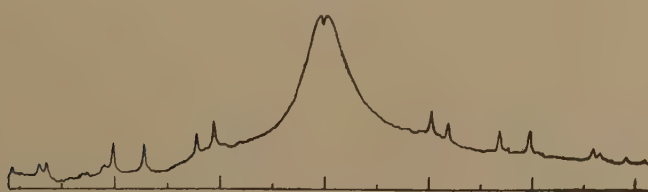
As already explained, the non-linearity of the valve characteristics introduces a distinction between D^* and D . The curves A, B, C, D of fig. 3 show the relation of D^* to D for filament currents of V_1 of 0.70, 0.67, 0.63 and 0.60 A. respectively. Curve A is linear up to densities of rather more than 0.8, and can be used when it is required to plot contours

Fig. 3.



of fairly strong lines, or to measure intensities. The slope of this line is about $\frac{1}{4}$, and when used in this way the microphotometer may be thought of as reducing the effective γ of the plate by a factor of 4. Increasing the filament current of V_1 above 0.70 A. changes the form of the D^* — D curve very little, and when used in this manner the instrument is extremely stable. Fig 4 is a plot of measurements made on a Debye-

Fig. 4.



Scherrer electron-diffraction photograph of etched gold-foil under these conditions. The large divisions along the base of the figure represent cms. The small reproduction cannot show all the details of the original figure, which was a plot of 600 readings. One hour was required to make all these measurements, and the drift in the microammeter during this period was only 1 per cent. of full-scale. Thus the time required to

make such a comprehensive and detailed survey of a plate is not excessive. The positions of the maxima on this figure could be determined to only 25μ , but the line regions can, of course, be investigated in greater detail if desired. We have not had occasion to make any detailed study of line intensities, but there is every indication that the microphotometer is suitable for this purpose.

For the precise measurement of line position it is better to adjust the instrument so that the relation between D^* and D is of the form of curves B, C or D of fig. 3, the value of D at the middle of the steep portion of the curve being that at the centre of the line. This has the effect of reducing the apparent width of the line by a factor of at least two. This adjustment is made as follows. The plate carriage is set just beyond the edge of the line, and the filament current of V_1 is so adjusted that the change in the reading of G on switching off the lamp is a maximum. The filament current is now slightly increased to its optimum value; when the filament current is adjusted in this way, we get the greatest change in the reading of G in passing through the line, and at the same time the smallest width of line. We have investigated the characteristics of the AT40, using a Kipp and Zonen type Zc galvanometer, and find that under the above conditions the diode is just saturated; thus a small change in the photo-cell current produces a large change in potential across it. Therefore the potential of the grid of V_2 rises in passing through the line from almost earth potential to slightly above the filament potential. Further rise of potential is prevented by grid current in V_2 , and it is apparent that the magnitude of this rise in grid potential is practically equal to the bias voltage produced across R_1 . This bias voltage is 2 v., which is sufficient to give reasonably large currents through G , while at the same time avoiding serious curvature of the characteristic of V_2 . If a D^*-D curve were to be plotted, it would then be found to be of the general form of C of fig. 3. The lower portion corresponds to the grid potential of V_2 being limited by grid current, and, therefore, changing little with photo-cell current; the middle portion corresponds to the rapid change of voltage across the saturated diode V_1 with small change of photo-cell current; the upper portion corresponds to a smaller rate of change of potential across V_1 for currents below saturation value.

For the determination of the centre of a line, it is unnecessary to plot a complete contour of the line. In the case of a symmetrical line, it suffices to take about twelve readings of the microammeter for equally spaced settings of the plate carriage arranged roughly symmetrically about the centre of the line. The true centre of the line is then the "centre of gravity" of all the readings. A commercial adding machine can be used for calculations of this type.

The position of a line of moderate density of total width 500μ can be determined to within 1μ provided the line is fairly symmetrical, while even very weak lines of the same width, which are invisible in an ordinary measuring microscope, can be measured to within 2 or 3μ . If the filament current of V_1 is not set exactly at the optimum value, no error

is introduced into the measurements of the position of the line, though the apparent width is, of course, increased. For the success of this method of measuring the position of a line it is obviously of paramount importance that the density fluctuations due to graininess of the plate shall be a minimum. We have found Ilford Special Lantern or Thin Film Half-tone plates to be suitable when developed with a borax fine-grain developer, the latter plate giving the finer grain.

We use this microphotometer mainly for precision measurements of line position by the method just indicated. On the few occasions when we wish to make intensity measurements, we use a D^*-D curve of type A of fig. 3, by setting the filament current of V_1 to just over 0.7 A. If, however, extensive intensity measurements are to be made, it would be better to use a D^*-D curve of the form E of fig. 3, as this gives larger deflections of the microammeter for a given density change, permitting greater accuracy in the measurement of the weaker lines. A curve of this form can be obtained if the filament current of V_1 is 0.90 A. and if the grid is disconnected from the anode and given a bias voltage of -0.75 v. to earth. This voltage can readily be obtained by connecting the grid to the negative terminal of the 150-v. battery and interposing a resistance of 0.8 ohm between this terminal and earth.

Acknowledgment.

It is a pleasure to express our thanks to Professor J. A. Crowther for the facilities placed at our disposal for the work, and for the advice and encouragement we have received.

XXIII. *The Development of the Trichromatic Theory of Colour Vision.*

By Prof. W. PEDDIE.

(Phil. Mag. ser. 7, vol. xxxiii. p. 559, Aug. 1942.)

A CORRECTION.

Correction is required in consequence of errors of sign in the three equations immediately below fig. 2, p. 568.

In all the equations on pp. 568, 569, with the exception of the expression for E in line 14, p. 569,—the sign of each σ which has two suffixes must be reversed. So the expression for E_2' should be

$$E_2' = E_1 \frac{1\sigma_2 - p(\sigma_2 + {}_2\sigma_3 + {}_2\sigma_1)}{\sigma_1 + {}_1\sigma_2 + {}_1\sigma_3 - p \cdot {}_1\sigma_2},$$

where $p = E_2/E_1$.

The condition for the two cortical stimuli having opposite signs is then

$$p < {}_1\sigma_2/(\sigma_2 + {}_2\sigma_3 + {}_2\sigma_1).$$

The following three lines should then read—"The second cortical stimulus then will be negative in this case. The third cortical stimulus is negative always in this case, while the first is always positive."

On p. 569, line 13 from foot, "indicates" should read "may indicate"; and on line 10 from foot, "negative" should read "small valued."

The first sentence at the top of p. 569 should be altered to "It is important to note that a cortical stimulus may be either positive or negative relatively to the primary stimulus," and in line 4, *necessarily* should be italicised.

XXIV. *New Contributions to Interferometry.—Part I.*

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

WITH reference to Dr. Tolansky's paper on this subject in the Phil. Mag. for August 1943, I wish to be permitted to draw the attention of your readers to a paper by myself and Mr. V. S. Rajagopalan in the 'Journal of the Optical Society of America' for October 1939 and to quote therefrom the description given on page 413 of essentially the same phenomenon. "It is worthy of remark that the Haidinger interferences in a transparent plate may also be exhibited either in reflection or in transmission using a point source of light. . . . A small aperture in an opaque screen closely backed by a mercury lamp is the origin of a divergent pencil in the path of which the transparent plate is placed. The reflected system of rings is then formed on the rear surface of the screen containing the illuminated aperture. The transmitted system of rings may be similarly observed when a diverging beam after a passage through the plate falls on a viewing screen. Though these arrangements are in certain respects less satisfactory than those in which extended light source is employed, they are nevertheless of interest as showing that such a source is not indispensable for the observation of Haidinger's rings." The geometric theory of this method of observing interferences and its extension to the case of curved plates are dealt with pretty fully in the same paper. It is scarcely necessary to add that the interferences seen in a Fabry-Perot Etalon are the same as the Haidinger rings of a parallel plate except that their sharpness is improved by the silvering of the surfaces.

Yours faithfully,

C. V. RAMAN.

Indian Institute of Science,
Malleswaram Post, Bangalore.
27th November, 1943.

XXV. *New Contributions to Interferometry.*

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

IN reading the article by S. Tolansky, "New Contributions to Interferometry, Part I. New Non-Localized Interference Fringes," *Phil. Mag.* (7) xxxiv. pp. 555–565, August 1943, I was reminded of an article: Robert Pohl, "Ein Neuer Versuch zur Interferenz divergierender Lichtbündel," 'Physikalische Zeitschrift,' vol. xli. pp. 498–499, November 1940.

Pohl used a mercury arc, sheet of mica approximately 0.04 mm. thick in place of a Fabry-Perot interferometer and reflected light instead of transmitted light, which Tolansky also considers.

Because of the small path difference Pohl's arrangement does not give narrow fringes nor high resolution and obviously is not suited for precise measurement.

Aside from this difference and the effect of double refraction and dispersion in the mica the theory of the two experiments is the same.

As both Pohl and Tolansky mention, their arrangements are admirably adapted to lecture demonstration. Pohl's article reproduces a photograph in which the whole projection wall of the lecture room in the Göttingen physics laboratory is covered with circular interference bands. Successive projection of the two arrangements would make an excellent demonstration of the increase of resolution with increasing path difference.

Yours faithfully,

L. B. TUCKERMAN,
Asst. Chief, Mechanics and
Sound Division.

National Bureau of Standards,
Washington.
Nov. 18, 1943.

XXVI. *New Contributions to Interferometry.—Part I. A Reply.*

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

THE notes by Sir C. V. Raman and Dr. L. B. Tuckerman which refer to my paper with the above title, call for comment. I shall discuss first the note by Raman. It seems to me that he completely disregards the essential distinguishing feature of my paper, namely, my insistence upon the importance of multiple-beam interference and my treatment of the effects thereof.

That an unsilvered parallel plate (having *only two effective interfering beams*) would lead to the formation of the circular rings described by Raman and Rajagopalan could have been inferred from the analogous well-known *circular fringes* formed by the Billet split lens arrangement (the Meslin modification reported in 1893). The optical conditions are closely similar in that in both cases there are two coherent point sources in line behind each other on the optic axis.

What is by no means immediately obvious is that a point source used with *multiple beams* would inevitably produce non-localized fringes which have the valuable characteristic sharpened distribution of the Fabry-Perot fringes and therefore exhibit a corresponding high resolving power. Indeed, as my analysis shows, this is by no means always the case, certain conditions requiring to be satisfied before very sharp rings appear. I have virtually taken it for granted that *two beams* would form circular rings, and my whole paper is devoted to the influence of *multiple beams*, which feature, as I repeat, Raman disregards above.

I must also certainly take strong objection to Raman's last sentence in which he closely identifies Haidinger rings with Fabry-Perot fringes in saying that the silvering merely "improves" the sharpness of the latter. This is very much of an understatement indeed, if the proper silverings are used (*it might stand with poor quality mirrors of low reflecting coefficient*); for whilst Haidinger rings are now of a purely didactic interest only, the enormous power of the highly sharpened distinctive Fabry-Perot fringes as an extremely important research tool in metrology, optics, nuclear spectroscopy, etc., is too well known to require emphasis.

I have pointed out possible serious research applications of the non-localized *multiple-beam* fringes, since they exhibit that very characteristic intensity distribution which makes the Fabry-Perot fringes so valuable. The *two-beam* fringes described by Raman and Rajagopalan simply cannot be compared with the fringes I have described in respect to brilliance, precision, resolution and general usefulness.

Finally, is it not a fact that Sir C. V. Raman is technically in error by identifying either his fringes or mine, since both are *non-localized*, with Haidinger rings of equal inclination, since these are at *infinity*? The fact that both are circular does not constitute identity. I have indicated many minor distinctions too.

Concerning the note by Dr. Tuckerman, I am much indebted to him for drawing my attention to the paper by Robert Pohl in 'Physikalische Zeitschrift' for Nov. 1940. Unfortunately, since the outbreak of the war we have failed to receive *any* German periodicals at Manchester and I was thus quite unaware of this interesting publication, which actually appeared about a year before I began my experiments. As Dr. Tuckerman so correctly points out, although there are a number of minor distinctions, the really important essential difference between Pohl's fringes and those I describe is the very high resolving power characteristic of mine (of the order of half a million in one case). It is just this very point that I have strongly emphasized, namely, the striking combination of enormous

dispersion, together with high resolution, despite the optical simplicity. I would add, however, that the high resolving power is not only due to the increased path difference mentioned by Dr. Tuckerman, but is equally to be attributed to the extreme fringe sharpening consequent upon the employment of many multiple beams. Perhaps a useful analogy would be to compare the two-beam interference with the multiple-beam interference as one can compare a grating having only 2 lines with one having 30 or 40 lines (the equivalent "equal" beams).

Yours faithfully,

S. TOLANSKY.

Physics Department,
Manchester University.
April 1st, 1944-

XXVII. *Notices respecting New Books.*

A German Physics Reader. By J. E. CALTHROP. [Pp. 83.] (London and Toronto: Wm. Heinemann, Ltd., 1943. Price 7s. 6d.)

MR. CALTHROP has provided the student of physics with a book of which it may be said with literal truth that it "fulfils a long-felt want."

The usual type of science reader covers a wide range of topics and the space devoted to any one branch of science is necessarily limited. The student of physics, for example, does not obtain anything like the amount of graded reading necessary in tackling so difficult a language as German.

This book is divided into three sections. The first section consists of eleven chapters devoted to topics treated in historical fashion, and varying from Brownian motion to Radioactivity. Each chapter is accompanied by an almost literal translation. The second section is made up of some thirty carefully chosen chapters, each selected to illustrate an important physical topic, and each chapter has its own vocabulary. The third section is devoted to German questions, of the "translate and comment on the following passage" type, set in recent London University Examinations at the Intermediate, General and Special Honours Grades.

The student who has conscientiously worked through this series of exercises will have acquired a sound knowledge of the construction of simple German sentences together with a very useful working vocabulary; he will, moreover, have reviewed, rapidly but not necessarily superficially, a number of topics of fundamental importance. The book can be unreservedly commended.

A. F.

Reports on Progress in Physics.—Vol. IX. (1942–1943). [Pp. iv+349.] (The Physical Society: London, 1943. Price to non-Fellows 25s. net.)

RATHER more than a quarter of this volume is taken up by reports of the Grassiot Committee of The Royal Society on Atmospheric Physics and Chemistry. In his introduction to this section, Professor Chapman points out that the physical and chemical conditions of the atmosphere depend largely on the sun's radiation and its interaction with the atmosphere. The nine reports which follow deal with the ultra-violet spectrum of the sun (about

which little is known with certainty), the absorption and emission spectrum of the atmosphere, the photochemical processes possibly occurring in the upper air and their counterparts in the laboratory. While in the absence of adequate data much of the work is speculative and inferential, the reports do most clearly state the present position and problems and indicate the purpose and scope of future work.

After a review of recent developments in Solar Physics, the volume ranges over a wide field including rubber, monolayers, counter tubes, field emission of electrons, crystal dynamics of crystals as revealed by X-ray scattering, the physical search for oil, sound (a general review of progress), and finally, the physics of painting.

The Society is once more to be congratulated on obtaining informative, critical, and up-to-date contributions from active workers in their respective fields. In a period when, of necessity, much exciting work must remain unpublished and much laid aside, it is no mean achievement to have given such evidence as this volume does of the continued vigour and scope of Physical Science.

J. E.

*[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]*

XXVIII. *The Influence of Crystal Size on the Absorption Factor as applied to Debye-Scherrer Diffraction Patterns.*

By A. TAYLOR, M.Sc., Ph.D., F.Inst.P.*

[Received September 15, 1943.]

SYNOPSIS.

The absorption factor derived by the methods of Claassen and Bradley for a cylindrical crystal is applied to the case of a spherical crystal, for which a set of absorption factor tables is constructed. It is then shown that the absorption factor A for a cylindrical powder specimen consists of two parts, a micro-absorption factor τ corresponding to the individual reflecting crystallites which are assumed to be spherical, and α , the macro-absorption factor for the specimen as a whole. The effects of crystallite aggregation, size and dilution upon A are discussed. Each constituent in a mechanical mixture must be given its own absorption factor on account of the different values of the micro-absorption factors which can only be neglected when the dimensions of the crystallites fall below 10^{-5} cm. The influence of the micro-absorption factor on the evaluation of the amplitude of atomic vibration is also briefly discussed.

I. *Introduction.*

WHEN X-rays fall upon a powder specimen and give rise to the well-known Debye-Scherrer type of diffraction spectrum, an appreciable fraction of the radiation is absorbed in the specimen, and consequently there is a diminution in the intensities of the spectrum lines. The extent of the absorption depends on μ , the linear absorption coefficient of the material, and upon the dimensions of the specimen which govern the mass of material to be traversed by the X-rays. If an X-ray beam of intensity I_0 per square centimetre falls upon the surface of a medium of absorption coefficient μ , the intensity I in the beam after traversing a distance z cm. is given by the exponential relation

$$I = I_0 e^{-\mu z} \quad (1)$$

Since it is mass of material encountered which is effective, and not the length of path, it is customary to use the mass absorption coefficient $\frac{\mu}{\rho}$ which is listed in tables. Equation (1) may therefore be written

$$I = I_0 e^{-\frac{\mu}{\rho} \cdot \rho z} \quad (2)$$

* Communicated by the Author.

The mass absorption coefficient $\frac{\mu}{\rho}$ is a complex function of the atomic number of the absorbing element and of the wave-length of the radiation. It reaches sharply defined peak values at certain wave-lengths and then falls steeply at what are termed "absorption edges." If the beam traverses a medium made up of the mass fractions a of element A, b of element B, etc., then the mean mass-absorption factor of the medium will be

$$\begin{aligned} \left(\frac{\mu}{\rho}\right)_m &= \left(\frac{\mu}{\rho}\right)_A \cdot a + \left(\frac{\mu}{\rho}\right)_B \cdot b + \dots + \left(\frac{\mu}{\rho}\right)_I \cdot i + \dots \\ &= \Sigma \left(\frac{\mu}{\rho}\right)_I \cdot i. \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (3) \end{aligned}$$

In the usual Debye-Scherrer arrangements, the specimen is shaped from a mass of crystal fragments into a tiny cylinder some 0.03 to 0.10 cm. in diameter with a height of 0.2 to 0.4 cm. bathed in the incident radiation. Calculating the precise amount of absorption in such a specimen is greatly complicated by the variations in path length which the various rays in the incident and reflected beams must traverse in the circular cross-section. By means of a rigorous mathematical analysis based on Claassen's method ⁽¹⁾, A. J. Bradley ⁽²⁾ was able to derive expressions for evaluating the absorption in cylindrical Debye-Scherrer specimens.

Bradley's expressions are applicable to the zero-order layer-line spectra from cylindrically shaped single crystals rotated about the cylinder axis in addition to powder specimens. In the latter case, the assumption is made that the specimen is a homogeneous mixture having an average mass absorption coefficient $\left(\frac{\mu}{\rho}\right)_m$ and neglects the effect of the finite size of the individual crystals. In the analysis given below, the Bradley method will first of all be applied to small single crystals and the results extended to cover the case of a cylindrical powder specimen, in which an allowance will be made for the effects of grain size. It is of interest to note that K. Schäfer ⁽³⁾ has also observed that grain size has an appreciable influence on the value of the absorption factor, and attempts to make an allowance for it in the case of flat specimens.

II. Diffraction by a Single Crystal.

Consider a very small single crystal completely bathed in a parallel beam of unpolarized monochromatic radiation of wave-length λ , and rotated about a given zone axis with uniform angular velocity $\frac{d\theta}{dt}$. A reflection will flash out as the crystal rotates through the narrow angular range $\theta_0 - \epsilon$ to $\theta_0 + \epsilon$, where θ_0 is the Bragg angle given by the relation $2d \sin \theta_0 = n\lambda$. If we let $R(\theta)$ be the reflecting power of the set of crystal planes at the angle θ , and if I_0 is the intensity of the incident beam, then, in the absence of absorption, the intensity of the diffracted ray will be $I_0 R(\theta)$.

The total energy I reflected into the spectrum line will therefore be

$$I = \int_{\theta_0 - \epsilon}^{\theta_0 + \epsilon} I_0 R(\theta) \cdot d\theta$$

$$= I_0 QV, \quad \dots \quad (4)$$

where V is the volume of the crystal and

$$Q = \frac{\lambda^3}{\sin 2\theta_0} \left(\frac{Ne^2}{mc^2} \right)^2 \frac{1 + \cos^2 2\theta_0}{2} \{F(hkl)\}^2,$$

e/m = ratio of electronic charge to mass,

N = number of unit cells in the crystal per unit volume,

$F(hkl)$ = the structure factor, and

$$\frac{1 + \cos^2 2\theta_0}{2} = \text{the polarization factor.}$$

In the presence of absorption, the value of I/I_0 is reduced by a factor α , which is termed the absorption factor. We must therefore write (4) in the form

$$I/I_0 = \alpha \cdot QV. \quad \dots \quad (5)$$

α is a function of the mass absorption coefficient, $\frac{\mu}{\rho}$, and the dimensions of the crystal.

III. Evaluation of the Absorption Factor α for a Cylindrical Crystal of Radius r .

In the absence of absorption, the integrated reflection I/I_0 is proportional to V, the volume of crystal immersed in the X-ray beam. For a cylindrical crystal, the integrated intensity is therefore proportional to $\pi r^2 h$, where h is the height of crystal in the beam and r its radius. We may divide the crystal into elemental prisms of height h and cross-sectional area $d\sigma$. The volume of these extremely small elements is therefore $dV = h d\sigma$.

In the presence of absorption, the contribution of these small elements of volume to the integrated intensity is no longer QdV but $QdV e^{-\mu z}$, where μ is the linear absorption coefficient of the crystalline material and z the total length of path through the crystal before and after reflection (see figure). The volume dV can be made so small that absorption within it can be considered to be negligible.

To evaluate the total strength of the reflection we must integrate the expression $Qe^{-\mu z} dV$ throughout the volume of the specimen, and the ratio

$$\alpha = \frac{\iiint Qe^{-\mu z} dV}{QV}$$

$$= \frac{1}{V} \iiint e^{-\mu z} dV \quad \dots \quad (6)$$

instant, and the diameter of the halo, part only of which is used for intensity measurements; and $F(hkl)$ the structure factor.

The constant is a function of e/mc , λ and V the total volume of the reflecting crystals.

We determine A in a manner similar to α for the cylindrical single crystal; only now the specimen consists of a uniform mixture of powder particles of small size and Canada Balsam adhesive. We shall, for the sake of simplicity, assume the particles to be single crystals, spherical in shape and with a radius small compared with the radius of the cylinder. Nevertheless, the absorption within these small spheres will be considered finite, and we shall term this effect "micro-crystalline" or "micro-absorption."

Consider now a narrow pencil of X-rays, PA, which forms part of the beam irradiating the specimen. On entering the specimen, the ray enters a uniform mixture of crystal particles and Canada Balsam, to which we can give an average mass absorption coefficient $\left(\frac{\mu}{\rho}\right)_m$. Since

this absorption is on the scale of the specimen, we shall term it the "macro-scale" or "macro-absorption." After travelling the distance p , it encounters the spherical crystal at B, which absorbs an appreciable part of the energy of the X-ray beam along a very short internal path in the process of reflecting the beam along the path BC of length q .

The integrated reflection from the small spherical crystal of radius a and volume v allowing for the micro-absorption is equal to τQv , where τ is the absorption factor for the sphere. This is still further reduced by the macro-absorption of the incident and reflected beams in the distance $p+q$. The contribution of the crystal to the intensity of the Debye-Scherrer halo given by the specimen as a whole is, therefore, cut down to

$$\tau Qve^{-\mu(p+q)} \quad \text{or} \quad \tau Qve^{-\left(\frac{\mu}{\rho}\right)_m \rho m^2}.$$

Ignoring for a moment the angular factor, planar factor, etc., the total contribution made by the spherical particles to the integrated reflection is

$$\Sigma \tau Qve^{-\left(\frac{\mu}{\rho}\right)_m \rho m^2} \quad \dots \quad (14)$$

Now in the absence of absorption, the total intensity would have been proportional to QV' where V' is the total volume of crystal powder in the specimen. The intensity of reflection from the specimen as a whole is therefore cut down by absorption in the ratio

$$A = \frac{\Sigma \tau Qve^{-\left(\frac{\mu}{\rho}\right)_m \rho m^2}}{QV'} \quad \dots \quad (15)$$

Now if dV is an element of volume of the specimen, it will include an average mixture of particle and Canada Balsam. If V' is the total volume of crystal powder in the specimen and V the volume of the entire

specimen, which consists of balsam and crystal, we may replace v by $\left(\frac{V'}{V}\right) \cdot dV$ and V' by $V \left(\frac{V'}{V}\right)$. Equation (15) then becomes

$$A = \frac{\sum \tau Q \left(\frac{V'}{V}\right) dV e^{-\left(\frac{\mu}{\rho}\right)_m \rho m^2}}{Q V \left(\frac{V'}{V}\right)} \\ = \frac{\sum \tau e^{-\left(\frac{\mu}{\rho}\right)_m \rho m^2} dV}{V} \dots \dots \dots (16)$$

We can now carry out the summation through the volume of the specimen. Rewriting equation (16), we have

$$A = \frac{\tau}{V} \iiint e^{-\left(\frac{\mu}{\rho}\right)_m \rho m^2} dV.$$

Therefore, according to equation (6),

$$A = \tau \alpha. \dots \dots \dots (17)$$

In other words, to determine the absorption factor A for a powder specimen, we must first of all find the micro-absorption factor τ for the individual crystallites with absorption coefficient $\frac{\mu}{\rho}$ and the macro-absorption factor for the specimen with an average mass absorption coefficient $\left(\frac{\mu}{\rho}\right)_m$. The product of the micro- and macro-absorption factors then yields the true absorption factor, A , of the specimen.

VIII. Effects of Crystallite Size, Aggregation and Dilution.

So far we have regarded the crystals as discrete spheres of equal radius. In the majority of cases, the powder fragments will consist of crystallite aggregations, and the dimensions of the individual crystallites will possess some form of statistical distribution.

Consider a single fragment of the powder. Since the orientations of the component crystallites will be perfectly random, only one crystallite at a time will be in a position to reflect as the specimen rotates in the X-ray beam. The micro-absorption factor, τ , refers to this one reflecting crystallite. The remaining bulk of the powder fragment absorbs without contributing to the reflection at that particular instant, and its contribution to the total absorption is accounted for in the macro-absorption factor, α , which refers to the specimen as a whole.

On the other hand, the distribution of crystallite size must affect the ultimate value of τ . The smaller crystals are the more efficient reflectors on account of their τ -factors, and any size distribution will be skewed in their favour when considering their contribution to the total intensity of the spectrum line. Should it be required to refine the calculations further to take this into account, the values of τ given in Table II. may be modified by multiplying them by the weight fraction of crystallites

having the radius a_c and summing over the entire distribution. This will be no easy matter, as crystallite size measurements have not yet attained such a degree of precision. For the present we must be content with an "effective crystallite radius" in computing the value of τ .

If the specimen consists of a mixture of powder and Canada Balsam, the τ term is unaffected by the extent of the dilution, but α is very sensitive to it. As the specimen is made more and more dilute, α tends more and more to the absorption factor for Canada Balsam. If the crystallites have a high value of $\frac{\mu}{\rho}$, the controlling term in the absorption factor of the super-dilute specimen will clearly be τ . On the other hand, for densely packed specimens with a high value of $\left(\frac{\mu}{\rho}\right)_m$, the more important factor will be the macro-absorption factor α . Thus the neglect of including the micro-absorption factor in the customary method of computing A leads to very inaccurate results with dilute specimens.

IX. *The Absorption Factors for a Cylindrical Specimen with Two or more Constituent Powders.*

If the specimen consists of a mixture of the substances P, Q, R... to whose crystallites we can apply the micro-absorption factors $\tau_P, \tau_Q, \tau_R \dots$, then the absorption factors for the individual overlapping Debye-Scherrer patterns will be

$$A_P = \tau_P \alpha_{P+Q+R+\dots},$$

$$A_Q = \tau_Q \alpha_{P+Q+R+\dots},$$

$$A_R = \tau_R \alpha_{P+Q+R+\dots},$$

where $\alpha_{P+Q+R+\dots}$ is the macro-absorption factor for the specimen as a whole. That is, we have to apply a different absorption factor to the Debye-Scherrer spectrum of each discrete substance which registers a pattern on the photograph. Only when the crystallites are of the order of 10^{-5} cm. in size do the micro-absorption factors approach unity (or 100 per cent., as in the tables), in which case the absorption factors $A_P, A_Q \dots$, etc. converge upon a single average value $\alpha_{P+Q+R+\dots}$.

Let us see how the neglect of including appropriate τ -factors will influence the result of a chemical analysis made by X-ray diffraction methods. Writing the observed intensities in the form $I = I_0' A$, we see from equations (13) and (4) that I_0' is directly proportional to the volume and, therefore, to the mass of diffracting crystals in the specimen. I_0' is thus *inversely* proportional to A and, therefore, to $\tau\alpha$. Suppose we have powders of copper and aluminium with the same crystallite radius $1/1000$ cm. mixed in the ratio $m : 1$ by weight. Let us further suppose that we are comparing the ratio of Debye-Scherrer reflections yielded by these two substances with CuK_α radiation in the region of a Bragg angle $\theta = 45^\circ$. Let α be the macro-absorption factor for the cylindrical specimen as a whole and $\tau_{\text{Cu}}, \tau_{\text{Al}}$ the micro-absorption factors of the

copper and aluminium crystallites respectively; then, since $I'_{0(\text{Cu})}$ and $I'_{0(\text{Al})}$ are proportional to the masses of the ingredients, we have

$$\frac{\text{mass of Cu}}{\text{mass of Al}} = \frac{m}{1} = \text{constant} \cdot \frac{I'_{0(\text{Cu})}}{I'_{0(\text{Al})}},$$

$$= K \frac{\tau_{\text{Al}} \alpha}{\tau_{\text{Cu}} \alpha},$$

i. e.

$$m = K \frac{\tau_{\text{Al}}}{\tau_{\text{Cu}}}.$$

If we neglect the micro-absorption factors, we obtain an erroneous value m' given by

$$m' = K.$$

The error in the determination will therefore be

$$\left(\frac{m - m'}{m} \right) \cdot 100 = \left(1 - \frac{\tau_{\text{Cu}}}{\tau_{\text{Al}}} \right) \cdot 100 \text{ weight per cent.}$$

For CuK_α radiation, $\theta = 45^\circ$ and $a = 10^{-3}$ cm., we have the following data :—

	$\frac{\mu}{\rho}$	ρ	$\mu a = \left(\frac{\mu}{\rho} \right) \rho a$	100τ
Cu.....	52.7	8.9	0.46	52.4
Al.....	48.7	2.7	0.13	83.2

Thus neglect to include the τ -factors leads to an apparent relative deficiency in copper of

$$\left(1 - \frac{52.4}{83.2} \right) \cdot 100 = 37 \text{ per cent.}$$

(or an apparent excess of aluminium of 59 per cent.). If we reduce the radii of the crystallites of both powders to 10^{-4} cm., this error drops to 7 per cent., while a radius of 10^{-5} cm. leaves a residual error of only 0.5 per cent. Schäfer (*loc. cit.*) has already demonstrated this effect in the case of aluminium-iron powder mixtures.

The extension of the problem from mixed powders to two-phase alloys becomes obvious, for a two-phase alloy yields an X-ray diffraction pattern which consists of the overlapping patterns from each phase. These patterns should have relative intensities which are in the same ratio as the weight fractions of the two constituents. We now see that this cannot be so, for, on account of the τ -factors, one phase will absorb more radiation than the other, and so its pattern will be recorded with too low an intensity, and may, in fact, be entirely suppressed ⁽⁴⁾.

In most alloy work care is taken to give a heat treatment to the powder so as to obtain sharp end-doublets. If the crystallite dimensions

are to be in the region of 10^{-5} cm. to minimize the τ -factors, then an appreciable broadening, β , of the doublets will be observed, as can be computed from the Scherrer formula,

$$\beta = \frac{1.0 \lambda}{\epsilon \cos \theta}$$

where λ is the wave-length of the radiation, ϵ the crystallite size and θ the Bragg angle. For truly sharp reflections, the size of the crystals must be above 10^{-4} cm. in order to avoid the Scherrer broadening. If we wish to retain sharp end-doublets, we have to work in a region where the effects of crystallite size make appreciable contributions to the absorption factors of each phase. It is probably the neglect of the micro-absorption factors which has led X-ray workers to differ in many instances from users of the microscope when determining the positions of phase boundaries from the relative intensities of the diffraction patterns.

X. *Effects of Crystallite Size upon the Debye-Waller Factor.*

In structural work in which the Debye-Scherrer method is used, we attempt to correlate the experimentally determined values of $F(hkl)$ of equation (13) with those calculated theoretically on the basis of an assumed atomic arrangement. Now equation (13) includes the two terms, A , the absorption factor, and the Debye-Waller temperature factor, $e^{-2B \frac{\sin^2 \theta}{\lambda^2}}$, the former increasing with $\sin^2 \theta$ while the latter decreases with it. To some extent, then, they tend to cancel each other out.

If we compute an incorrect value for A by ignoring the τ -factor, this error can be minimized by suitably adjusting the value of B in the Debye-Waller factor. This makes little or no difference to the final atomic arrangement, but it does mean that an error has been introduced into the determination of the amplitude of atomic vibration.

Conclusions.

The Claassen-Bradley derivation of the absorption factor for cylindrical crystals has been discussed and applied to the case of spherical crystals. It has been shown that in a cylindrical Debye-Scherrer powder specimen the absorption within each individual reflecting crystallite yielding the micro-absorption factor, τ , must be included in addition to the macro-absorption factor, α , which refers to the specimen as a whole. Failure to include the τ -factor leads to incorrect values of the intensities of the spectrum lines, which, in turn, adversely influence calculations on the quantities of phases in polyphase alloys. Only when the crystallite size falls below 10^{-5} cm. can errors from this source be neglected. This, unfortunately, is a region in which line-broadening occurs, and is valueless for alloy work. Some correction must therefore be made for the existing crystallite size in the practical range which lies above 10^{-4} cm.

TABLE I.

Absorption Factors 100α when $\mu r < 5$ for Cylindrical Specimens.
(A. J. Bradley, Proc. Phys. Soc. xlvii. p. 879, 1935.)

$\sin^2 \theta$	0	0.1464	0.3290	0.5000	0.6710	0.8536	1.00
$\begin{array}{c} \theta \\ \mu r \end{array}$	0°	$22\frac{1}{2}^\circ$	35°	45°	55°	$67\frac{1}{2}^\circ$	90°
0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.1	84.70	84.80	—	84.9	—	85.0	85.1
0.2	71.20	71.60	—	71.9	—	72.4	72.9
0.3	60.00	60.60	—	61.4	—	62.7	63.5
0.4	51.00	51.70	—	52.7	—	54.5	55.6
0.5	43.50	44.20	—	45.8	—	47.8	49.0
0.6	36.90	37.80	—	39.8	—	42.3	43.6
0.7	31.40	32.40	—	34.8	—	37.8	39.3
0.8	26.80	27.85	—	30.5	—	33.7	35.6
0.9	23.00	24.10	—	27.1	—	30.5	32.4
1.0	19.77	20.95	—	24.2	—	27.85	29.5
1.1	16.98	18.28	—	21.70	—	25.50	27.15
1.2	14.59	16.00	—	19.54	—	23.50	25.10
1.3	12.56	14.03	—	17.70	—	21.70	23.35
1.4	10.84	12.33	—	16.11	—	20.10	21.80
1.5	9.38	10.91	—	14.69	—	18.66	20.50
1.6	8.11	9.73	—	13.52	—	17.46	19.32
1.7	7.10	8.71	—	12.47	—	16.41	18.24
1.8	6.15	7.80	—	11.54	—	15.42	17.30
1.9	5.37	7.02	—	10.74	—	14.59	16.44
2.0	4.71	6.35	—	10.05	—	13.84	15.67
2.1	4.16	5.79	—	9.44	—	13.15	14.93
2.2	3.67	5.31	—	8.89	—	12.50	14.26
2.3	3.24	4.86	—	8.38	—	11.89	13.65
2.4	2.865	4.47	—	7.91	—	11.35	13.09
2.5	2.55	4.12	—	7.50	—	10.86	12.56
2.6	2.27	3.82	—	7.11	—	10.40	12.11
2.7	2.02	3.55	—	6.75	—	9.98	11.67
2.8	1.803	3.30	—	6.41	—	9.62	11.27
2.9	1.607	3.08	—	6.10	—	9.25	10.89
3.0	1.436	2.885	4.43	5.82	7.24	8.89	10.54
3.1	1.288	2.705	4.22	5.58	6.97	8.57	10.21
3.2	1.159	2.55	4.02	5.35	6.72	8.30	9.90
3.3	1.049	2.415	3.84	5.14	6.48	8.04	9.61
3.4	0.955	2.29	3.68	4.95	6.26	7.78	9.33
3.5	0.871	2.17	3.525	4.77	6.05	7.55	9.06
3.6	0.796	2.06	3.385	4.60	5.85	7.33	8.81

TABLE I. (cont.)

$\sin^2 \theta$	0	0.1464	0.3290	0.5000	0.6710	0.8536	1.00
$\begin{array}{c} \theta \\ \mu r \end{array}$	0°	22 $\frac{1}{2}$ °	35°	45°	55°	67 $\frac{1}{2}$ °	90°
3.7	0.729	1.968	3.255	4.44	5.66	7.11	8.58
3.8	0.670	1.875	3.14	4.29	5.49	6.92	8.36
3.9	0.617	1.787	3.03	4.15	5.33	6.73	8.15
4.0	0.568	1.706	2.925	4.02	5.17	6.53	7.94
4.1	0.525	1.629	2.825	3.89	5.02	6.35	7.74
4.2	0.488	1.563	2.73	3.77	4.88	6.18	7.55
4.3	0.453	1.500	2.645	3.66	4.75	6.02	7.38
4.4	0.420	1.445	2.555	3.56	4.62	5.87	7.21
4.5	0.391	1.390	2.485	3.47	4.50	5.73	7.05
4.6	0.364	1.343	2.41	3.38	4.39	5.61	6.89
4.7	0.340	1.300	2.34	3.29	4.28	5.47	6.75
4.8	0.316	1.259	2.275	3.21	4.18	5.35	6.61
4.9	0.2945	1.222	2.21	3.13	4.08	5.25	6.47
5.0	0.2755	1.189	2.155	3.05	3.99	5.14	6.35

TABLE II.

Absorption Factors 100τ for Spherical Particles.

$\sin^2 \theta$	0	0.1464	0.5000	0.8536	1.0000
$\begin{array}{c} \theta \\ \mu a \end{array}$	0°	22 $\frac{1}{2}$ °	45°	67 $\frac{1}{2}$ °	90°
0.00	100.0	100.0	100.0	100.0	100.0
0.01	98.6	98.6	98.6	98.6	98.6
0.02	97.2	97.2	97.3	97.3	97.3
0.03	95.9	95.9	96.0	96.0	96.0
0.04	94.5	94.5	94.6	94.6	94.7
0.05	93.2	93.2	93.3	93.3	93.4
0.06	91.8	91.8	91.9	92.0	92.1
0.07	90.5	90.5	90.6	90.6	90.8
0.08	89.2	89.2	89.3	89.4	89.5
0.09	87.8	87.8	87.9	88.0	88.2
0.10	86.5	86.5	86.6	86.8	87.0
0.20	74.2	75.0	75.3	76.0	76.0
0.30	64.0	65.1	65.3	65.8	67.1
0.40	54.5	56.6	56.9	58.0	58.7
0.50	46.8	48.9	49.6	51.6	52.9

TABLE II. (*cont.*)

$\sin^2 \theta$	0	0.1464	0.5000	0.8536	1.0000
$\begin{array}{c} \theta \\ \mu a \end{array}$	0°	22½°	45°	67½°	90°
0.60	41.0	42.2	43.6	46.0	47.7
0.70	36.0	37.2	38.7	41.8	43.4
0.80	31.5	32.6	34.7	37.9	39.9
0.90	27.6	28.7	31.1	34.5	36.4
1.00	23.9	25.1	28.2	31.8	33.4
1.10	20.9	22.2	25.5	29.1	30.9
1.20	18.4	19.7	23.2	26.8	28.7
1.30	16.2	16.6	21.2	24.8	26.8
1.40	14.5	16.0	19.4	23.2	25.0
1.50	12.9	14.3	17.9	21.8	23.5
1.60	11.4	12.9	16.5	20.7	22.1
1.70	10.1	11.8	15.2	19.6	20.9
1.80	9.0	10.6	14.1	18.5	19.8
1.90	8.0	9.6	13.1	17.6	18.8
2.00	7.1	8.7	12.3	16.8	18.0
2.10	6.2	7.8	11.4	16.0	17.2
2.20	5.6	7.1	10.8	15.3	16.4
2.30	5.0	6.5	10.2	14.5	15.7
2.40	4.5	6.0	9.6	13.8	15.2
2.50	4.0	5.5	9.1	13.2	14.6
2.60	3.6	5.1	8.7	12.6	14.1
2.70	3.3	4.8	8.3	12.0	13.6
2.80	3.0	4.6	7.9	11.4	13.0
2.90	2.8	4.3	7.6	11.0	12.6
3.00	2.6	4.1	7.3	10.6	12.2
3.50	1.7	2.8	5.9	8.9	10.5
4.00	1.2	2.1	5.0	7.81	9.3
4.50	0.9	1.9	4.2	7.10	8.3
5.00	0.6	1.7	3.7	6.0	7.4

The non-inclusion of the micro-crystalline absorption factor in structure work leads to modifications in the Debye-Waller temperature factor which in turn yields erroneous values of the amplitude of atomic vibration.

Acknowledgment.

The author wishes to thank the English Electric Co., Ltd., for permission to publish this paper.

References.

- (1) A. Claassen, *Phil. Mag.* ix. p. 57 (1930). F. C. Blake, *Rev. Mod. Phys.* v. p. 169 (1933).
- (2) A. J. Bradley, *Proc. Phys. Soc.* xlvii. p. 879 (1935).
- (3) K. Schäfer, *Zeits. für Physik*, lxxxvi. p. 738 (1932).
- (4) M. Gayler, Discussion on the Determination of Equilibrium Diagrams by X-ray Methods, *J. Inst. Met.* lxix. p. 1 (1943).

XXIX. *New Contributions to Interferometry.* Part IV.—*The Ghost Images and Scatter Rings of the Fabry-Perot Interferometer and their Effects in Hyperfine Structure Observations.*

By S. TOLANSKY, Ph.D., D.I.C., Manchester University*.

[Received November 14, 1943.]

[Plate IV.]

SUMMARY.

The origin of the ghost images due to the reflections at the unsilvered faces of a Fabry-Perot interferometer is discussed. The effects of these images upon hyperfine structure investigations is considered. The way in which the particular type of crossing with a spectrograph which is used influences the behaviour and effects of the ghost images is examined. The effects of scattered light from the intense beam back-reflected from the first silvered surface are considered. It is shown that with an injudiciously chosen crossing arrangement, both the ghosts and the scatter fringes can seriously impair observations either on weak multiplet lines or on weak hyperfine structure components. It is shown that all defects due to ghosts and scatter are eliminated by employing the well-known external parallel beam crossing mount, providing this is modified by the introduction of an auxiliary slit at the focus of the first parallelizing lens. There is no loss of light with this arrangement. The mounting described is the only one which can be safely used under all conditions and ranges of intensity.

Section I.—SECONDARY IMAGES.

Introduction.

THE Fabry-Perot interferometer has found extensive employment within the last twenty years for the examination of the hyperfine structures of spectrum lines. In the earlier studies the use of the instrument was generally restricted to the investigation of the few strong lines in individual spectra. In the larger number of the cases examined the

* Communicated by the Author,

intensities of the individual constituent components of the hyperfine structure pattern were comparable. With the general progress in the study of hyperfine structure it has become necessary to examine *weak* spectrum lines in the close neighbourhood of strong lines and also the detection of *weak components* within an individual hyperfine structure pattern has assumed increasing importance. Such weak components may either represent relatively rare isotope lines, or may belong to weak transitions in a nuclear spin multiplet. The importance of detecting rare isotope lines is obvious, but no less important is the accurate measurement of the positions of spin multiplet components, for the discovery of the existence of nuclear electric quadrupole moment and its effects upon the interval rule makes the accurate determination of the positions of weak components frequently a decisive factor.

Both the examination of a weak spectrum line close to a strong line in the same source, and the detection of weak components within a pattern containing one or more *much* stronger components, involve experimental difficulties which have so far not been considered in the literature on the use of the Fabry-Perot interferometer. Three main factors are involved namely,

- (1) the reflecting coefficient of the metallic coatings ;
- (2) the influence of the ghost images arising from reflections from the unsilvered faces of the interferometer plates ;
- (3) scatter effects.

Only the first factor, the influence of the reflecting coefficient on the performance of the instrument, has up to now been treated. This aspect has been so exhaustively examined that little further remains to be added. It is the object of this paper to examine the influences of factors (2) and (3) above.

The ghost images arising from reflection at the unsilvered faces are of course absent when an all-glass interferometer is used (a plane parallel plate of glass silvered on both sides). For many reasons the all-glass interferometer has *not* been generally adopted, the air-gap type offering many advantages. The discussion following refers to the air-gap type in general use, but in addition the conclusions reached about scatter effects apply equally well to the all-glass type.

In present-day practice the Fabry-Perot interferometer is usually crossed with a prism spectrograph, being placed either inside or outside of the latter. It has not been formerly recognized that the particular mode of crossing adopted has an important bearing on the influence of the ghost images and the scattering. This will be treated in detail later.

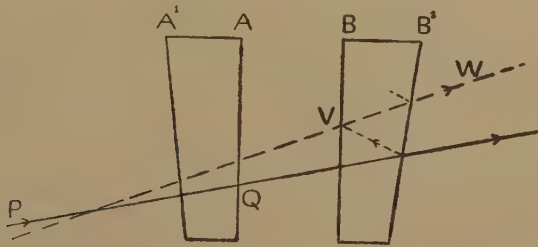
The Origin of the Ghost Images.

Although the ghost images are familiar to most users of the instrument, the detailed manner in which they arise has not been treated before. The two faces of each of the plates of an air-gap interferometer are not made plane parallel to prevent the formation of the complex fringes

arising from four plane parallel faces. Each plate has a slight wedge angle, and as 1 minute of arc suffices to prevent the formation of Fabry-Perot fringes between silvered and unsilvered surfaces, this value has been adopted by some makers. As will be shown later, *an angle of 10 minutes of arc is decidedly advantageous because of the influence of the ghost images, as distinct from spurious fringes.*

According to the crossing arrangement adopted, the fringes of any line will be accompanied by either one or two secondary ghost images displaced from the principal image by an amount depending on the wedge angle, and upon the direction of the wedge apex. Each secondary image has an intensity about 1/30th that of the principal image. The origin of these images has usually been briefly dismissed as being due to reflection at the unsilvered surfaces. This is in fact true, but as it happens, it has not been realized that the two secondary ghost images are formed in distinctly different ways, are usually different in appearance, and can have quite different influences on the spectrum under examination. To distinguish them they will be called (a) *the first-plate secondary image* and (b) *the second-plate secondary image*. The first-plate is that which *first* receives the light, *i. e.* that *nearer* the source.

Fig. 1.



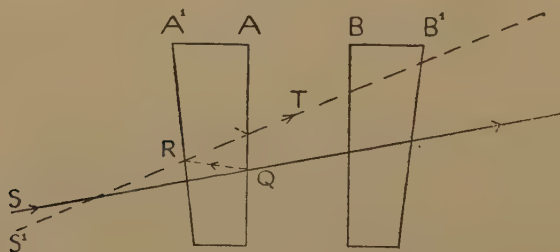
The second-plate secondary image appears with all forms of mounting. Fig. 1 shows how it arises. Let it be assumed that the reflecting coefficient of the silvered surfaces is of the order of 0.90 (the actual value is immaterial). AB and A'B' are the silvered and unsilvered faces of the first and second plates respectively. Since refraction can be disregarded, an incident ray in the direction PQ emerges as a group of parallel rays of diminishing intensity. The collection of these by a lens leads to the projection of the fringes at the focal plane. Some 4 per cent. of each ray in this bundle is reflected at the glass surface B', and of this 90 per cent. is reflected at the silvered surface B in a direction parallel to VW, so that there emerges in the latter direction a bundle of rays identical with that parallel to PQ but of some $3\frac{1}{2}$ per cent. intensity of the latter, and making an angle with it depending on the wedge angle between B and B'. Hence at the focal plane of the collecting lens, in addition to the principal system PQ, there is a secondary system VW about 1/30th as strong. Because of the angle between VW and PQ *the*

centre of the secondary rings does not coincide with that of the principal fringes.

That the second-plate secondary system does arise in the manner described above is proved by the simple fact that rotation of the second plate about the optic axis causes the secondary system to move as a whole in a circle about the centre of the principal system.

The first-plate secondary image is formed in a distinctly different way. It is only separately seen when the light which falls on the interferometer is parallel light coming from a slit (or its equivalent—a line source). In particular therefore it appears when the interferometer is placed inside a spectrograph, in the parallel beam, between the collimator lens and the prism, a mounting very frequently adopted. The mode of formation is shown in fig. 2. The incident light from the slit S , rendered parallel by a lens, falls upon the silvered face A . Consider a single ray SQ . Some 90 per cent. is reflected back along the direction QR , and of this 4 per cent. is reflected from the glass surface A' in the direction RT . In effect, therefore, there exists a virtual image S' of the point S (which is on the

Fig. 2.



slit), displaced from the latter by an amount proportional to the wedge angle between A and A' . This holds for each point on the slit, hence *there exists a virtual slit image 1/30th as intense as the true slit.*

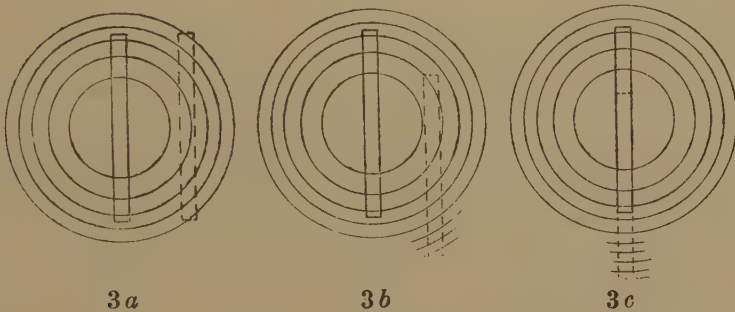
All light incident on the interferometer contributes to but *one* fringe system, and the effect of introducing the weak rays in the direction $S'T$ is to lead to the formation of weak rings of higher order number than those represented by the direction SQ . These weak rings belong to the same system, *have the same centre* and are merely part of the principal fringes. If therefore the slit be replaced by an extended source the only effect of the reflected beams $S'T$ is to contribute some more light into outer rings and no separate secondary system appears. The same is true if, instead of a parallel incident beam, convergent light is projected on to the interferometer.

When a slit is used, it is clear that a secondary *image* is formed. When an extended source is employed a complete circular ring system is projected, but the effect of a slit is to isolate a diametral section of the rings, since the angles of incidence are limited by the slit dimensions. The virtual image formed by the reflected light supplies beams at angles

other than those contributing to the principal image. It is as if an image of the slit were moved by the action of the wedge and placed over some portion of the ring system other than diametrically. This leads clearly to the appearance of a weak secondary *concentric* system.

The position occupied by the first-plate secondary image depends upon the orientation of the wedge apex of the first plate. This is illustrated in fig. 3. The slit selects the principal image (heavy lines) from the circular rings. The secondary image (dotted lines) appears at the positions shown in 3 *a*, 3 *b*, 3 *c*, when the wedge apex is placed respectively horizontal, at 45° to the horizontal, and vertical. *In no case does this particular secondary image confuse the principal image.* It is to be noted, however, that in case 3 *c* the intensities of the fringes in the overlap region are some $3\frac{1}{2}$ per cent. too great, because of the superposition of the secondary image. This fact has been completely overlooked in former publications dealing with intensity measurements.

Fig. 3.



It must be emphasized that whereas the second-plate secondary image (which will be considered later) overlaps with and confuses the principal image, this first-plate secondary image does not. The latter may, and frequently does, overlap and confuse the principal fringes of *other neighbouring lines* unless special precautions are adopted in a manner to be described.

The Influence of the Form of Crossing with a Spectrograph.

Three different methods have been developed for the crossing of a Fabry-Perot interferometer with a prism spectrograph. These can be called conveniently,

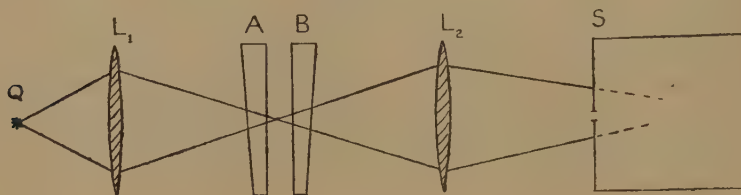
- (a) the external converging beam mounting,
- (b) the internal parallel beam mounting,
- (c) the external parallel beam mounting.

The effect of the secondary images depends upon the mounting employed. These will be considered below in turn.

(a) *The External Converging Beam Mounting.*

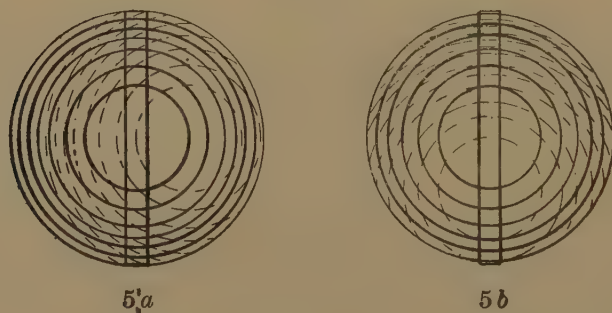
This is shown in fig. 4. The lens L_1 projects an image of the source Q on to the interferometer AB . The fringes, at infinity, are brought to a focus at S , the spectograph slit by L_2 . Since the image of Q and the fringes are not simultaneously in focus, all the incident light is not usefully employed. In any case the beams of high convergence lead to rings of very high order-number and these are not required.

Fig. 4.



With this arrangement there is *no first-plate secondary image*, the back reflected light leading to a $3\frac{1}{2}$ per cent. increase in intensity of the principal image. The second-plate secondary image is projected on to the slit as a *complete ring system*, but the centre is displaced from that of the principal rings. If ϕ is the wedge angle of the second plate, the linear displacement D of the centre of the second secondary image from the centre of the principal image is given by $D=2\phi l$ where l is the focal length of the lens L_2 . Taking l to be 1 metre (and shorter focal lengths are very frequently used), D is equal to 0.55 millimetre if ϕ is 1 minute of arc. I have adopted $\phi=10$ minutes of arc in my own instruments

Fig. 5.



giving $D=5.5$ millimetres. Even so there is partial overlapping in the manner shown in figs. 5 *a* and 5 *b*.

In fig. 5 *b* the wedge apex of the second plate has been set vertically. The ghost image (dotted) appears within the principal image (heavy lines) and the fringes of this can easily be mistaken for weak true components. They can be differentiated from the principal fringes by being out of step with these, since the centres do not coincide. This out of step

effect is not easy to detect when ϕ is small, and especially when the ring diameters are large because of a small plate separation. Even if unambiguously identified, these secondary fringes can cause serious disturbances. Weak components in the principal pattern may be completely masked in particular cases. In others the influence of the secondary image may be considerable even if the exposure is insufficient to make them separately visible. For they can add their light in an entirely arbitrary manner unsymmetrically on to fringes which are even appreciably strong and cause serious spurious displacement of the optical centre of gravity of the affected fringe. This is an extremely important matter in those investigations concerned with effects due to nuclear quadrupole moment.

It is quite clear, too, that apart from errors in position, grave errors can easily be made in evaluating intensities of *weak* components if the pattern under examination is being microphotometered. *Errors of 100 per cent. and more can easily arise in numerous typical hyperfine structure patterns.*

In fig. 5 *a* the wedge apex of the second plate has been set horizontal. In most respects this is more favourable than the position shown in fig. 5 *b*. The secondary fringes cross the principal rings at angles approaching 90° near to the centre of the latter, the angle falling off rapidly as one moves from the centre of the system. These crossed secondary fringes cannot be mistaken as false components, yet their influence is still considerable in specific cases. They can affect the apparent positions of true weak components and are particularly disastrous if these components are being examined with a microphotometer for intensities, since on the whole they occupy a considerable fraction of the slit area and their contribution to the background intensity varies from order to order. This effect must be emphasized since it has been disregarded in the literature.

Thus there is no satisfactory way for eliminating the second-plate secondary system when using the external converging beam mounting. This form of crossing is therefore unsuited to the accurate study of either the displacements or the intensities of the *weak* components of a hyperfine structure pattern. In particular, if isotope abundance ratios are being obtained, the abundances of the weaker isotopes will be suspect.

(b) *The Internal Parallel Beam Mounting.*

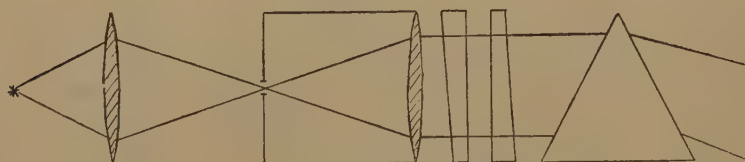
This mounting is much favoured, the arrangement being shown in fig. 6. The interferometer is placed in the parallel beam between the collimator lens and the prism (there are attendant disadvantages in placing it between prism and camera lens). Both first-plate and second-plate secondary images are formed. Since the slit selects a section of the ring system, each slit image is accompanied by two secondary slit images, the centres of which are displaced from the centre of the principal slit image by $2 l' \phi$ where l' is the focal length of the camera lens.

Both the first-plate and the second-plate secondary images are now

restricted to the slit dimensions (in the previous mounting discussed the second-plate secondary image forms a *complete* circular ring system). By selecting suitable directions for the wedges both images can be thrown clear from the principal image. *All confusion within the principal image is thus eliminated.*

It is advisable to set the two wedges in opposing directions otherwise the combination of the two wedge plates is equivalent to the introduction

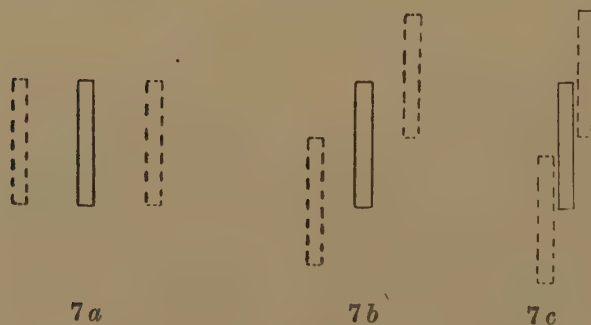
Fig. 6.



of a small angle prism into the spectrograph. With a wedge of $10'$ of arc this is enough to displace the spectrum sufficiently to affect the focus if the camera has a large tilt.

The positions of the secondary slit images (dotted) relative to the principal slit image are shown in fig. 7 for typical settings with the wedges with each apex horizontal (7 *a*), at 45° (7 *b*) and close to the vertical (7 *c*). All three arrangements separate the secondary images from the principal image, but it is to be noted that in 7 *a* and 7 *b* the ghosts can, in specific cases, superpose on neighbouring lines, and if the

Fig. 7.



latter are weak, can entirely swamp them. Thus the 7 *c* arrangement is frequently the most advantageous. With a sufficiently large wedge angle, it is clear that enough flexibility exists to permit of placing the secondary images so as to avoid falling on to any particular close neighbouring line.

It is of course clear that if the secondary ghost images are themselves sufficiently intense they will themselves produce ghosts. Thus tertiary and even quaternary ghost images can often be seen. The intensities

of successive ghost images are in the ratio $1 : 1/30 : 1/900 : 1/27,000 \dots$ so that extreme over-exposure of the principal image is necessary before higher order ghosts appear. Nevertheless they are surprisingly frequently seen on plates. Pl. IV. shows the principal image accompanied by one secondary, tertiary and quaternary ghost (the second set has been thrown down out of the picture).

It can be concluded that the internal parallel beam mounting can successfully eliminate any confusion within a given line due to secondary and higher order ghost images. But to avoid the confusion of neighbouring lines by the ghost image of a close-by strong line a sufficiently large value for the wedge angle must be adopted, and care must be taken to place the ghost images correctly to avoid the lines in question.

Fig. 8.

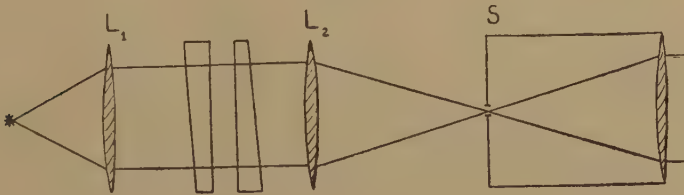
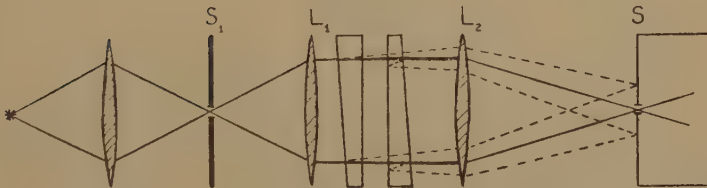


Fig. 9.



(c) The External Parallel Beam Mounting.

A number of well-known makes of spectrograph are built mechanically on lines which do not permit the introduction of an interferometer into the parallel beam. The mounting then usually adopted is the external parallel beam mount shown in fig. 8. With this arrangement both the image of the source and the fringes are simultaneously in focus on the spectrograph slit S . The light economy is therefore at its best. It is clear that the secondary ghost images are identical with those in the external converging beam mounting and all the defects of this arrangement also apply. Yet despite this either the one or the other arrangement *must* be adopted with many existing spectrographs. Since the importance of the secondary images has not been discussed before, no remedy has, of course, been yet suggested. By means of the simple device shown in fig. 9, all confusion effects of the ghost images can be entirely eliminated *without any loss in light in the principal image*.

As shown in plan in fig. 9, an auxiliary wide vertical slit S_1 is introduced at the focus of the lens L_1 . This is placed either directly in front of the source, or if this is not practicable, an image of the source is projected on to the slit S_1 . Both first-plate and second-plate secondary images are now separately formed (dotted). The wedge angles of the interferometer plates are set opposing, in a horizontal direction, and as a result both the ghost images are displaced sideways and do not enter the spectrograph slit by all.

Since no dispersion has yet taken place *all* the ghost images belonging to *all* the spectrum lines are automatically thrown to the sides. No ghost images at all appear on the photographic plate and, in addition to having eliminated confusion *within* a line, the superposition effects on neighbouring lines are also eliminated.

Thus it can be concluded that, as regards the elimination of secondary ghost images (and higher orders), the external parallel beam mounting *when used in conjunction with an appropriate auxiliary slit* is superior to other mountings.

It may be of interest to point out that at times some of the individual fringes in the secondary images are crossed by fine dark bands. It is probable that these are superposed Jamin or Brewster fringes.

In the succeeding section the effects due to scattered light will be considered.

Section II.—THE EFFECT OF SCATTERED LIGHT.

Effects due to scattered light have been disregarded in the literature, since they are only of importance when very strong and weak lines are both to be studied. The scatter effects are in some cases far more serious than those due to the ghost images. The form of the intensity distribution of the Fabry-Perot interferometer is such that the detection of hyperfine structure components of intensity less than 1 per cent. of the main component is a matter of considerable experimental difficulty. This is because the minimum between adjacent maxima does not go down to zero. The minima with reflecting coefficients of 0.80, 0.85, 0.90 have respectively intensities which are $1\frac{1}{2}$, $\frac{2}{3}$ and $\frac{1}{4}$ per cent. of the maxima. To this must be added the effects of scatter, and it will be shown that in unfavourable cases scatter effects are so drastic that not only weak components, but indeed complete line patterns can be obliterated entirely.

It happens frequently in many spectra that lines exist, not far apart, with intensity ratios of several thousands to one. For example, in many hyperfine structure problems helium is used as a carrier gas, and conditions of the discharge are often such that the helium lines are exceedingly bright compared with some of the weak lines under investigation, particularly if these involve unfavourable transitions. In such cases, scatter effects, if not guarded against, can entirely prevent observation being made on all the components, both weak and strong, of the weaker lines,

even if several hundred angströms distant from the intense line responsible for the scatter troubles.

The nature of the scatter defect depends entirely on the form of crossing adopted, and each will be briefly discussed in turn.

The source of the scattered light is largely the 90 per cent. of the incident beam which is reflected back from the first silvered surface. As already pointed out, 4 per cent. of this constitutes the first-plate secondary image which appears as a definite slit image pattern. The remaining considerable amount of light passes back to the lens which projects the beam on to the interferometer. This is variously the parallelizing lens, the collimator lens of the spectrograph, or, in the converging beam mount, the converging lens. Some 4 per cent. is returned back to the interferometer from each face of this lens. Hence, depending on lens design, at least 7 per cent. of the incident light is returned to the interferometer.

Owing to the curvatures of the lens faces, and any tilt on the interferometer, this light is not returned in a parallel beam but is in general irregularly reflected, supplying beams of wide angles of incidence. This leads to the formation of *complete Fabry-Perot rings*. It is as if the scattered light came from a large extended source, and not from a slit.

Since the 7 per cent. of back reflected light is distributed effectively over a much larger area than the slit image, the intensity of the scatter rings is considerably *less* than 7 per cent. of the principal image, yet in spite of this, the effects can be serious. The seriousness depends entirely on the mounting adopted for crossing, and these will be considered in turn.

(a) *The External Converging Beam Mount.*

Any scattered light has *no ill effects at all* and does not influence definition. The spectrograph slit cuts out all scatter rings, and as with the first-plate secondary image, the effect of scatter is to produce a slight (almost negligible) enhancement in the intensity of the principal image.

(b) *The Internal Parallel Beam Mounting.*

It has been shown already that, from the point of view of secondary ghost images, the internal parallel beam mounting is very little inferior to the external parallel beam mounting used with an auxiliary slit. This cannot be said of scatter effects which profoundly differ in the two arrangements.

The scatter effects can be most serious when using the internal parallel beam mounting. In addition to the secondary images on the plate there is superposed a complete ring system due to scattered light. The plate holder is usually tilted, as a result of which the rings are usually only in focus in the region in which the principal image occurs. In general, therefore, a diffuse slightly distorted ring system appears, but it

can be seen that the rings are true continuations of the fringes in the principal image.

The effect is shown very clearly in Pl. IV., in which the exposure has been chosen to show up the scatter rings produced by the yellow helium line, when helium has been used as a gas-discharge carrier in a hollow cathode-discharge tube giving out a metal spectrum difficult to excite, and therefore weak. The images marked (2), (3), (4) respectively are the secondary, tertiary and quaternary ghost images of the principal image, and it follows that the latter is over-exposed a matter of some 20,000 to 30,000 times. The rings are broad because of the natural broad width of the helium lines.

There is, of course, a considerable halo due to the excessive exposure, although backed plates were used. Apart from this, the scatter rings can be clearly seen, extending over a wide region and completely obscure a number of spectrum lines. It is important to note that the scatter rings extend far beyond the region of normal halo due to over-exposure. In Pl. IV. the scatter effect is obvious to the eye, yet it is clear that even if the scatter rings were reduced in intensity by a factor of 20 they would still have an effect on both the intensities and the apparent location of fringes. Clearly, then, if there exists a line, even no stronger than 1500 times the line under observation, and even if this strong line is some considerable distance from the weak line, there is a distinct danger of serious error arising.

This mounting is therefore not suited to the study of a spectrum containing both very strong and weak lines, all of which are to be examined.

(c) *The External Parallel Beam Mount.*

As in the case of the external converging beam mount, the scattered light has *no effect at all* and does not impair definition. In both the external mounts the spectrometer slit automatically cuts out the scatter rings, which are, of course, there all the time. It is naturally essential to take care that the angular aperture of the beam projected into the spectograph is slightly less than that of the collimator lens, otherwise there will be the usual general background spectrograph scatter from the walls of the collimator tube. This is usually less serious than the scatter due to the back-reflected beam. If not eliminated, it just adds to the general intensity of the scatter rings.

The external parallel beam mount is free from scatter ring defects whether an auxiliary slit is used or not.

Conclusion.

Since the external parallel beam mounting, *when used with an auxiliary slit*, completely eliminates the secondary images, and as it also eliminates the scatter rings, *it is the only mounting which can safely be employed for all cases*. If, then, a spectrum line is accompanied by strong neighbouring lines, or if within any given line there is an extreme intensity

range in the components of the hyperfine structure pattern, then the mounting to be adopted is the well-known external parallel beam mounting, modified by the introduction of an auxilliary slit at the focus of the first lens. The restriction does not apply if the few strongest lines in a spectrum are the only ones to be studied, and if the hyperfine structure components have comparable intensities.

XXX. *Kinematic Relativity.*

By Professor W. WILSON, F.R.S.*

[Received October 4, 1943.]

THE subject of relativity has recently acquired a new interest, in consequence of the work of Professor E. A. Milne, some account of which has appeared in recent papers in the 'Philosophical Magazine' †. Milne has developed a completely new kind of relativity. This in itself constitutes a great service to theoretical physics, whether his theory is destined to endure or not. Kinematic relativity, as Milne calls it, rests on foundations which are, or would appear to be, essentially different from those of the older Einstein-Minkowski type of relativity. It claims to be a relativity of observers. I take it that "observer" means a purposeful agent engaged in inquiry concerning the physical world and, while Milne's observers are included within this category, or in one sufficiently closely related to it, it is not possible to identify them with the observers in a physical laboratory or an astronomical observatory. The rôle of these latter is practically reduced to a rational and purposeful construction and disposition of apparatus, which, in fact, takes the observations and may even, and indeed sometimes does, record everything to which the physical or astronomical observer attaches importance; so that the observer himself, or another person, not involved in the experimental activity, might arrive later and note the result. Having done the construction and planning, the work of the laboratory observer is finished, so far as observing is concerned. This latter is the work of a robot. The founders of the older relativity recognized this, at least instinctively and tacitly. The principle of relativity makes no reference to observers. It is described by Professor Milne, correctly as I think, in the words: "The principle of relativity is usually stated in the form that the laws of nature are invariant in form for any arbitrary trans-

* Communicated by the Author.

† E. A. Milne, *Phil. Mag.* vol. xxxiv. pp. 73, 82, 197, 235 and 246 (1943). All references are to these papers, except where other works or papers are indicated.

formation of co-ordinates"*. As contrasted with this, Milne's relativity is based on "transformations from observer to equivalent observer." "The important thing," he says, "is not transformations of co-ordinates, but transformations of observers from one observer to another observer."

While greatly admiring Milne's effort, I find difficulties in accepting his results—including his equation (4), which appears to be fundamental. I propose to discuss some of these difficulties, especially one of them, in the hope that Milne, or someone sufficiently versed in his theory, may resolve it, or that, alternatively, some fundamental weakness may be revealed. Indeed, I am venturing to suggest that I have found a fundamental weakness.

Milne's equation (4) may be written

$$\frac{1}{Y} \frac{dV}{dt} = -\frac{1}{X} (P - Vt), \quad (1)$$

in which

$$X = t^2 - P^2/c^2, \quad Y = 1 - V^2/c^2. \quad (1')$$

Equation (1) gives the acceleration of "a free particle at the position P , moving with the velocity V at the instant t (all as reckoned by O)."

It may be remarked parenthetically that Milne gives a more general equation than (1) (*vide* 'Relativity, Gravitation and World Structure,' p. 95), namely,

$$g = (P - Vt) \frac{Y}{X} G(\xi),$$

in which g means dV/dt and $G(\xi)$ is an invariant to which the value -1 is assigned for reasons which do not concern me here.

Equation (1) emerges as the solution of certain functional equations which appear on p. 94 of his 'Relativity, Gravitation and World Structure.' It appears to be fundamental in his theory, since his claim to create *dynamics* out of a purely *kinematical* situation by logical methods involving no empirical appeals would seem to rest on the soundness of his deduction of (1). The functional equations referred to, which are based on the cosmological principle, appear to me to be sound, though I have not yet *quite* convinced myself of this.

Now one solution of these functional equations is

$$\frac{dV}{dt} = 0. \quad (1'')$$

* 'Relativity, Gravitation and World Structure' (Oxford, 1935).

Einstein's original statement of the (special) principle of relativity, as translated by Perrett and Jeffery, was "... the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good" ("Zur Elektrodynamik bewegter Körper," *Annalen der Physik*, xvii. (1905)). This statement, like that given by Milne, is a positive one. Relativity is not based on negative statements such as: "there is no meaning in absolute velocity," or "there is no meaning in absolute acceleration."

They are, in fact, satisfied by (1''), and nothing, except empirical appeal, justifies us in setting it aside.

Leaving this on one side, I have a difficulty of another kind in accepting (1). I agree that X is an invariant and that $\mathbf{P}-\mathbf{V}t$ is the space part of a 4-vector. Consequently the right-hand side of (1) must be the space part of a 4-vector*; but I am forced to insist that this is not true for the left-hand side of the equation. In the identity

$$\frac{1}{Y^2} \frac{d\mathbf{V}}{dt} = \frac{1}{Y^2} \frac{d}{dt} \left(\frac{\mathbf{V}}{Y} \right), \quad \dots \quad (2)$$

the right-hand side is the space part of a 4-vector, as Professor Milne agrees (see his equation (6)). So that this must be true of $(1/Y^2)(d\mathbf{V}/dt)$. How, then, can the left-hand side of (1) be the space part of a 4-vector? Surely only if Y is an invariant, which, however, is not the case.

On the other hand, Milne's equation (6), which, I must agree, follows of necessity from his (4) (*i. e.* from equation (1) above), *preserves its form under a Lorentz transformation!* This equation is

$$\frac{1}{Y^2} \frac{d}{dt} \left(\frac{\mathbf{V}}{Y} \right) = -\frac{1}{X} \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y} \right), \quad \dots \quad (3)$$

in which

$$Z = t - \mathbf{PV}/c^2. \quad \dots \quad (3')$$

How has this real, or apparent, contradiction slipped in? Perhaps the following illustration will throw some light on this: Let \mathbf{P} and \mathbf{Q} be the space parts of two 4-vectors, so that

$$\mathbf{P} - \mathbf{Q} \quad \dots \quad (4)$$

is the space part of a 4-vector, while

$$(\mathbf{P} - \mathbf{Q})/Y \quad \dots \quad (5)$$

is *not* the space part of a 4-vector, since Y is not invariant. Apart from the invariant factor, $-1/X$, the right-hand side of (3) is, in fact, equal to $(\mathbf{P} - \mathbf{V}t)/Y$. Now

$$\frac{\mathbf{P} - \mathbf{Q}}{Y} = \mathbf{P} + \phi - \frac{\mathbf{Q}}{Y},$$

provided we mean by ϕ

$$\phi = \frac{\mathbf{PV}^2}{Yc^2}.$$

Therefore we have the identity

$$\frac{\mathbf{P} - \mathbf{Q}}{Y} = \mathbf{P} - \mathbf{Q} \left(\frac{1}{Y} - \frac{\mathbf{PV}^2}{Yc^2} \right). \quad \dots \quad (6)$$

We ask now: How is it possible for us to be lured into the belief that the right-hand side of (6) is the space part of a 4-vector? The \mathbf{P} and \mathbf{Q} are

* I am also leaving on one side the important question as to whether Milne has succeeded in deducing the Lorentz transformation.

4-vectors, so that if the whole expression is a 4-vector, the expression multiplying \mathbf{Q} must be invariant. Thus

$$\frac{1}{Y} - \frac{\mathbf{P}\mathbf{V}^2}{\mathbf{Q}Yc^2} = I,$$

where I is invariant. An invariant is not necessarily a constant. It must simply have the same value in all the systems of reference with which we are concerned: in this case in the systems of reference of *all* Milne's equivalent observers. There is one equivalent observer in whose co-ordinate system $\mathbf{V}=0$. In this system $I=1$, and consequently

$$\frac{1}{Y} - \frac{\mathbf{P}\mathbf{V}^2}{\mathbf{Q}Yc^2} = 1$$

in the systems of reference of all Professor Milne's equivalent observers. Thus \mathbf{P} and \mathbf{Q} must be equal to one another, and this must be true for *any* vectors \mathbf{P} and \mathbf{Q} . Equation (6) is of course an identity, and its right-hand side becomes identical with that of Milne's (6), *i. e.* with that of (3) above, when

$$\mathbf{Q} = \mathbf{V}t$$

(assuming we have multiplied both sides of (3) by the invariant X).

It appears then that $\mathbf{P} - \mathbf{V}t$ must vanish, and we must conclude that

$$\frac{d\mathbf{V}}{dt} = 0.$$

The free particles in Milne's world have, therefore, *constant velocities* and are not distinguishable from his fundamental particles. The fact is that equation (1), with $\mathbf{P} \neq \mathbf{V}t$, is not compatible with the Lorentz transformation, which certainly governs the changes from one equivalent observer to another. The free particles can apparently only be accelerated in a continuum with some other, *e. g.* Riemannian, kind of metric.

Equation (3) is very important, in Milne's theory, and it gives a strong impression of being an equation whose members represent the space parts of 4-vectors and, therefore, an equation which conserves its form under a transformation from one observer to another, as, indeed, Milne believes. Equation (1) is undoubtedly equivalent to it and, rather naturally, he infers that (1) "conserves its form under the transformation from O to any O' ." It is well, at the risk of some repetition, to examine it more closely. Its right-hand side can be written

$$-\frac{1}{X} \left\{ \mathbf{P} - \mathbf{V} \frac{(t - \mathbf{P}\mathbf{V}/c^2)}{(1 - \mathbf{V}^2/c^2)} \right\}.$$

The part

$$\mathbf{V} \frac{(t - \mathbf{P}\mathbf{V}/c^2)}{(1 - \mathbf{V}^2/c^2)}, \quad \dots \dots \dots (7)$$

if written in the form

$$\frac{\mathbf{V}}{Y^{\frac{1}{2}}} \cdot \frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} \left\{ \frac{1}{2} \left(t^2 - \frac{\mathbf{P}^2}{c^2} \right) \right\},$$

has seemingly all the credentials of the space part of a 4-vector. As Professor Milne points out, $\mathbf{V}/Y^{\frac{1}{2}}$ is indeed the space part of a 4-vector, $Y^{\frac{1}{2}}dt$ is invariant and so is $t^2 - \mathbf{P}^2/c^2$. When, however, we write (7) in the form

$$\mathbf{V}t \frac{(1 - \mathbf{P}\mathbf{V}/tc^2)}{(1 - \mathbf{V}^2/c^2)}, \quad \dots \dots \dots (7')$$

we see that the factor multiplying $\mathbf{V}t$ must be invariant, since $\mathbf{V}t$ is indubitably the space part of a 4-vector, being, indeed, the displacement of a particle moving with a velocity \mathbf{V} in the system of reference of some observer, O , during the time, t , as reckoned by O . Now if the factor multiplying $\mathbf{V}t$ in (7') is invariant, it follows necessarily that $\mathbf{P} = \mathbf{V}t$. So once again we arrive at the conclusion that Milne's equation (4) reduces to

$$\frac{d\mathbf{V}}{dt} = 0.$$

That there may be no doubt about this conclusion let us scrutinize (7') in still another way. Let us write

$$\mathbf{D} = \mathbf{V}t.$$

It is the displacement which would occur in the time t if the moving particle had the velocity \mathbf{V} , all in the reckoning of one of Milne's observers, O . Let us transform from O to some other equivalent observer, O' , moving in O 's system with some velocity \mathbf{U} . The appropriate Lorentz equation is

$$\mathbf{D} = \frac{1}{(1 - \mathbf{U}^2/c^2)^{\frac{1}{2}}} (\mathbf{D}' + \mathbf{U}t').$$

Should the particle in question be (momentarily it may be) at rest in the reckoning of O' , then $\mathbf{U} = \mathbf{V}$ and $\mathbf{D}' = 0$. Therefore

$$\mathbf{D} = \frac{1}{Y^{\frac{1}{2}}} \cdot \mathbf{V}t',$$

and consequently

$$t = t'/Y^{\frac{1}{2}}. \quad \dots \dots \dots (7'')$$

But the adoption of the Lorentz transformation necessitates that

$$t' = \frac{1}{Y^{\frac{1}{2}}} (t - \mathbf{P}\mathbf{V}/c^2),$$

and when we combine this with (7'') we find once again $\mathbf{P} = \mathbf{V}t$ and, by (1), $d\mathbf{V}/dt = 0$.

It follows from this that

$$Z/Y = t, \quad \dots \dots \dots (8)$$

$$X = t^2 Y, \quad \dots \dots \dots (8')$$

and the invariant function, ξ , defined by

$$\xi = Z^2/XY, \quad \dots \dots \dots (8'')$$

which Milne finds to be constant for free particles, now becomes unity.

The defining equations (9) and (9') on p. 89 in Milne's paper II. now reduce to

$$\frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(m \xi^{\frac{1}{2}} \frac{\mathbf{V}}{Y^{\frac{1}{2}}} \right) = \mathbf{F}, \quad (9)$$

$$\frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(m \xi^{\frac{1}{2}} \frac{c}{Y^{\frac{1}{2}}} \right) = \mathbf{F}_t, \quad (9')$$

since we now have

$$\mathbf{P} = \mathbf{V}Z/Y.$$

But $\xi^{\frac{1}{2}}$ is now unity for a free particle and has lost the special significance it possessed, while the validity of (1) or (3) was not questioned. There is now no reason to complicate the definition of force by retaining it—no more reason than there would be in introducing any arbitrary invariant function of \mathbf{V} . So that (9) and (9') become

$$\frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(m \frac{\mathbf{V}}{Y^{\frac{1}{2}}} \right) = \mathbf{F}, \quad (10)$$

$$\frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(m \frac{c}{Y^{\frac{1}{2}}} \right) = \mathbf{F}_t. \quad (10')$$

It should be mentioned that m is defined by Milne to be a constant and to have the same value for all his equivalent observers.

These equations, it will be noticed, are part of the fundament of the special theory of relativity, and $(\mathbf{F}, \mathbf{F}_t)$ has now become the Minkowskian force, though sometimes the time component is defined (and this is very helpful I think) by

$$\mathbf{F}_w = \sqrt{-1} \mathbf{F}_t, \quad (11)$$

since it enables us to write all four components in the same form.

The mass of the particle is

$$M = m/Y^{\frac{1}{2}}$$

or

$$M = m/(1 - \mathbf{V}^2/c^2)^{\frac{1}{2}}. \quad (12)$$

This is indeed the universally adopted definition of mass and, being a definition, there is no question of experimental verification unless, or until, we have some independent identification of \mathbf{F} .

If we multiply both sides of (10) and (10') by $Y^{\frac{1}{2}}$ we get expressions for the Newtonian force, namely,

$$\frac{d}{dt} (M\mathbf{V}) = Y^{\frac{1}{2}} \mathbf{F} = \mathbf{f}, \quad (13)$$

$$\frac{d}{dt} (Mc) = Y^{\frac{1}{2}} \mathbf{F}_t = \mathbf{f}_t, \quad (13')$$

and when, in the case of a charged particle, for example, in motion at right angles to a magnetic field, we adopt the identification

$$M\mathbf{V}^2/r = \mathbf{f} = H e \mathbf{V}, \quad (13'')$$

we find, as an experimental fact *, excellent agreement with formula (12).

* Established by the work of Bucherer and others,

on the recession of the spiral nebulae, to calculate M_0 and thus to find the value

$$\rho_0 = 10^{-27} \text{ gram cm.}^{-3}$$

for the mean density of matter in the universe, a value in reasonably good agreement with that which the (rather rough) observations yield. It must be frankly conceded that this is a remarkable result; but there are many instances in the history of physical science of correct formulæ emerging from unacceptable theories. It may suffice to mention that of Fresnel for the convection of light. Einstein in his 'Kosmologische Betrachtungen' arrived at certain formulæ from which I infer that

$$\gamma = \frac{\pi R c^2}{2 M},$$

in which M and R are respectively the mass and the radius of his universe, and it may be noted that, if we replace R by ct , we get

$$\gamma = \frac{\pi c^3 t}{2 M},$$

which differs from Milne's formula in a quite unimportant way, in so far as its agreement with observational results is concerned. It is, of course, a high-handed procedure to replace Einstein's R by ct ; but it does indicate that the adoption of the cosmological principle (which I do not criticize) in theories otherwise very different from that of Milne is almost certain to yield a formula differing little from (17).

The foundations on which Milne's theory is built up are in accord with what is known about the recession of the spiral nebulae. There is at least so much of empirical appeal in his theory at the very outset. He assumes a certain cosmological principle. This is hypothesis. The Lorentz equations, which he uses, are, in his theory, the mathematical expression of certain conventions. The only solution of his functional equations (which I believe to be correct inferences from the cosmological principle as he interprets it) is that expressed by (1'') above, *i. e.* the only solution which is compatible with the Lorentz transformation. This cuts away, in my view, the foundation of his theory of gravitation and his rational electrodynamics. As contrasted with Milne's relativity, the Einsteinian type of relativity provides a completely satisfactory theory of gravitational phenomena; satisfactory from the point of view of its logical and mathematical coherence as well as from the point of view of its extraordinary agreement with observational results. Indeed, more than this can be claimed for it. If my memory serves me, it was a development of Einstein's cosmological theory, by de Sitter, which first suggested the possibility of the recession of the spiral nebulae and which, indeed, led to a formula for the dependence of the velocity of recession on the distance, a formula subsequently confirmed by the observation and measurement of the Doppler displacements,

XXXI. *Solutions of some Problems in Viscous Flow.*

By GEORGE GREEN, D.Sc., Lecturer in Applied Physics in the
University of Glasgow*.

[Received September 8, 1943.]

PROBLEMS relating to viscous flow as a rule present considerable difficulty owing to the nature of the boundary conditions to be fulfilled. These problems also do not provide the scope for the application of conjugate functions which obtains in the case of the corresponding problems relating to the flow of non-viscous fluid. Apart from some well-known standard cases, relatively few solutions of problems in viscous flow have been recorded⁽¹⁾. In the present paper it is proposed to record a number of solutions representing viscous flow within certain specified boundaries: these solutions being obtained by a method involving the application of conjugate functions which seemed to be worthy of fuller investigation.

The motions to be considered are in two dimensions, and the assumptions made are that the motion is small and steady, and that the density of the fluid is constant. Under these conditions the hydrodynamical equations to be satisfied are

$$\frac{1}{\rho} \frac{\partial p}{\partial x} = \frac{\mu}{\rho} \nabla^2(u), \quad \dots \dots \dots (1)$$

$$\frac{1}{\rho} \frac{\partial p}{\partial y} = \frac{\mu}{\rho} \nabla^2(v), \quad \dots \dots \dots (2)$$

together with the equation of continuity

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0. \quad \dots \dots \dots (3)$$

If we now introduce ζ , the vorticity, defined by the equation

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \quad \dots \dots \dots (4)$$

then ζ represents twice the angular velocity of spin of the fluid element at point x, y in an anti-clockwise direction. From (1), (2) and (3) we obtain

$$\nabla^2(\zeta) = 0; \quad \nabla^2(p) = 0; \quad \dots \dots \dots (5)$$

together with the relations

$$\frac{1}{\mu} \frac{\partial p}{\partial x} = -\frac{\partial \zeta}{\partial y}, \quad \frac{1}{\mu} \frac{\partial p}{\partial y} = \frac{\partial \zeta}{\partial x}. \quad \dots \dots \dots (6)$$

* Communicated by the Author,

These prove that $\frac{p}{\mu}$ and ζ are conjugate functions connected by an equation of the type

$$\frac{p}{\mu} - i\zeta = f(x + iy), \quad (7)$$

where f represents any function of $(x + iy)$. In the method employed below we shall suppose $(x + iy)$ expressed in terms of $(\xi + i\eta)$ where ξ and η are variables chosen and defined to suit the special boundary form in each particular problem.

The stream function ψ for any fluid motion may be expressed either in (x, y) or (ξ, η) co-ordinates. Taking $d\psi$ to represent an increment of flow from left to right we have (in x, y co-ordinates),

$$d\psi = u dy - v dx, \quad (8)$$

which gives

$$u = \frac{\partial \psi}{\partial y}, \quad v = -\frac{\partial \psi}{\partial x} \quad (9)$$

These relations yield $\zeta = -\nabla^2(\psi)$; and equations (1) and (2) give then as equation to determine ψ ,

$$\nabla^4(\psi) = -\nabla^2(\zeta) = 0. \quad (10)$$

Each fluid motion problem involves the determination of ψ in accordance with this equation and at the same time to fulfil certain boundary conditions. The general solution of (10), representing a real value of ζ , may be written in the form

$$\zeta = \frac{1}{i} \{f(x + iy) - f(x - iy)\}, \quad (11)$$

where f is an arbitrary function: and the determination of ψ in any particular problem depends first upon the finding of the correct form for f , and thereafter upon the finding of the complete integral of the equation

$$\nabla^2 \psi = -\zeta. \quad (12)$$

A similar process may be employed to obtain a solution when ψ is expressed in curvilinear co-ordinates (ξ, η) . In this case we have velocity components at point (ξ, η) expressed by

$$u = \frac{1}{h_2} \frac{\partial \psi}{\partial \eta}, \quad v = -\frac{1}{h_1} \frac{\partial \psi}{\partial \xi}, \quad (13)$$

and the circulation round a very small rectangle ABCDA enclosing the point (ξ, η) whose sides are $AB = h_1 d\xi$ and $AD = h_2 d\eta$ respectively, is given by

$$\frac{1}{h_1 h_2} \left\{ \frac{\partial(v h_2)}{\partial \xi} - \frac{\partial(u h_1)}{\partial \eta} \right\} \cdot h_1 h_2 d\xi d\eta, \quad (14)$$

which is equal to $-\zeta \times \text{area of rectangle ABCDA}$. Hence it follows that

$$\frac{1}{h_1 h_2} \left\{ \frac{\partial^2 \psi}{d\xi^2} + \frac{\partial^2 \psi}{d\eta^2} \right\} = -\zeta, \dots (h_1 = h_2), \quad (15)$$

this being the equation in terms of (ξ, η) co-ordinates corresponding to equation (12) in (x, y) co-ordinates. We may represent ζ by

$$\zeta = \frac{1}{i} \{F(\xi + i\eta) - F(\xi - i\eta)\}, \quad \dots \dots \dots (16)$$

an equation equivalent to (11) in x, y co-ordinates; and the final solution in (ξ, η) co-ordinates requires the complete integral of

$$\nabla^2(\psi) = -h_1 h_2 \zeta, \quad \dots \dots \dots (17)$$

the equivalent to (12) above.

Since ζ may be expressed in terms of $(x + iy)$ or of $(\xi + i\eta)$, we may define the co-ordinates (ξ, η) by an equation of the type

$$x + iy = f(\xi + i\eta), \quad x - iy = f(\xi - i\eta). \quad \dots \dots \dots (18)$$

These yield

$$(dx^2 + dy^2) = f'(\xi + i\eta) f'(\xi - i\eta) \{d\xi^2 + d\eta^2\}, \quad \dots \dots \dots (19)$$

so that $h_1 = h_2 = h$, and we may write

$$h^2 = h_1 h_2 = f'(\xi + i\eta) f'(\xi - i\eta). \quad \dots \dots \dots (20)$$

Consider now the expression

$$\psi_P = [f(\xi + i\eta)F(\xi - i\eta) - f(\xi - i\eta)F(\xi + i\eta)]. \quad \dots \dots \dots (21)$$

It yields

$$\nabla^2(\psi_P) = 4[f'(\xi + i\eta)F'(\xi - i\eta) - f'(\xi - i\eta)F'(\xi + i\eta)]. \quad \dots \dots \dots (22)$$

$$= f'(\xi + i\eta) f'(\xi - i\eta) \cdot 4 \left[\frac{F'(\xi - i\eta)}{f'(\xi - i\eta)} - \frac{F'(\xi + i\eta)}{f'(\xi + i\eta)} \right]. \quad \dots \dots \dots (23)$$

$$= -h_1 h_2 \zeta, \quad \dots \dots \dots (17)$$

provided that ζ is represented by

$$\zeta = 4 \left[\frac{F'(\xi + i\eta)}{f'(\xi + i\eta)} - \frac{F'(\xi - i\eta)}{f'(\xi - i\eta)} \right]. \quad \dots \dots \dots (24)$$

On the same assumption it is obvious that (21) is the particular integral of (17) above. The complementary function is any solution of

$$\nabla^2(\psi) = 0,$$

so that ψ_c = any potential function of the co-ordinates (ξ, η) .

The complete solution of (17) may then be written in the form

$$\psi = \psi_P + \psi_c. \quad \dots \dots \dots (25)$$

From the above it becomes evident that the value of ψ in any particular case is immediately obtainable when the function F' , which appears in (23), has been determined. Any arbitrary function of $(\xi + i\eta)$ and $(\xi - i\eta)$ inserted in (23), leads to a solution satisfying the fundamental hydrodynamical equations and the continuity equation. The finding of the exact solution of any problem thus depends on the possibility of determining the function F' with the aid of the boundary conditions applicable to the problem.

Flow within Converging Plane Walls ⁽⁵⁾.

As an illustration of the application of this method let us examine the case represented by

$$x + iy = ce^{\alpha(\xi + i\eta)} = f(\xi + i\eta), \quad . \quad . \quad . \quad (26)$$

$$h = h_1 = h_2 = c\alpha e^{\alpha\xi}, \quad . \quad . \quad . \quad (27)$$

$$x = ce^{\alpha\xi} \cos \alpha\eta, \quad y = ce^{\alpha\xi} \sin \alpha\eta, \quad . \quad . \quad . \quad (28)$$

$$r = ce^{\alpha\xi}, \quad \frac{y}{x} = \tan \alpha\eta. \quad . \quad . \quad . \quad (29)$$

In accordance with (23), put

$$\zeta = 4 \left[\frac{F'(\xi + i\eta)}{c\alpha e^{\alpha(\xi + i\eta)}} - \frac{F'(\xi - i\eta)}{c\alpha e^{\alpha(\xi - i\eta)}} \right]; \quad . \quad . \quad . \quad (30)$$

and the equation to determine ψ becomes

$$\nabla^2 \psi = -c^2 \alpha^2 e^{2\alpha\xi} \zeta. \quad . \quad . \quad . \quad (31)$$

The surfaces, $\eta = \text{constant}$, are planes passing through the origin. The surfaces, $\xi = \text{constant}$, are cylinders with OZ as axis. We may take the planes $\eta = \pm \eta_0$ as boundaries. As a test to find F, we examine first the possibility that a solution may exist for which ψ depends on η only and the stream lines are radial lines passing through the origin. If we take

$$F'(\xi + i\eta) = \frac{1}{\zeta} e^{-\alpha(\xi + i\eta)}. \quad . \quad . \quad . \quad (32)$$

this condition is fulfilled and

$$\nabla^2 \psi = 8c\alpha \sin 2\alpha\eta \quad . \quad . \quad . \quad (33)$$

$$\text{and} \quad \psi_P = -2 \frac{c}{\alpha} \sin 2\alpha\eta. \quad . \quad . \quad . \quad (34)$$

For the potential function we have

$$\psi_0 = A\eta, \quad . \quad . \quad . \quad (35)$$

and the complete integral of (33) is

$$\psi = -\frac{2c}{\alpha} \sin 2\alpha\eta + A\eta. \quad . \quad . \quad . \quad (36)$$

The boundary condition being

$$\frac{1}{h} \frac{\partial \psi}{\partial \eta} = 0, \quad \text{at } \eta = \pm \eta_0,$$

we find that $A = +4c \cos 2\alpha\eta_0$, and

$$\psi = -2 \frac{c}{\alpha} \sin 2\alpha\eta + 4c \cos 2\alpha\eta_0 \cdot \eta. \quad . \quad . \quad . \quad (37)$$

This solution is applicable to conditions of flow at large distance from the origin. From it we obtain

$$\zeta = -\frac{8 \sin 2\alpha\eta}{c\alpha e^{2\alpha\xi}}, \quad p = \text{constant} - \frac{8\mu \cos 2\alpha\eta}{c\alpha e^{2\alpha\xi}} \quad . \quad . \quad . \quad (38)$$

and

$$u = \frac{1}{h} \frac{\partial \psi}{\partial \eta} = - \frac{4(\cos 2\alpha\eta - \cos 2\alpha\eta_0)}{\alpha e^{\alpha\xi}} \quad (39)$$

As this solution satisfies prescribed velocity conditions at the boundary it must represent the only type of steady motion possible within the region occupied by the fluid—in accordance with the theorem of Helmholtz⁽⁶⁾.

Flow within Hyperbolic Boundaries ^{(1), (3)}.

When the equation defining co-ordinates (ξ, η) is

$$x + iy = c \cosh(\xi + i\eta) = f(\xi + i\eta), \quad (40)$$

the curves, $\xi = \text{constant}$, are ellipses

$$\frac{x^2}{c^2 \cosh^2 \xi} + \frac{y^2}{c^2 \sinh^2 \xi} = 1, \quad (41)$$

and the curves, $\eta = \text{constant}$, are hyperbolas

$$\frac{x^2}{c^2 \cos^2 \eta} - \frac{y^2}{c^2 \sin^2 \eta} = 1. \quad (42)$$

The curves are confocals, with foci $x = \pm c$. The curve, $\eta = 0$, coincides with the axis of x in the two ranges $x > c$ and $x < -c$.

In this case we put

$$h_1 h_2 = h^2 = c^2 \sinh(\xi + i\eta) \sinh(\xi - i\eta) \quad (43)$$

and

$$\zeta = 4 \left[\frac{F'(\xi + i\eta)}{c \sinh(\xi + i\eta)} - \frac{F'(\xi - i\eta)}{c \sinh(\xi - i\eta)} \right]. \quad (44)$$

Again looking to the possibility of a solution for which ψ depends on η only, we take

$$F'(\xi + i\eta) = \frac{-c}{4i} \cosh(\xi + i\eta), \quad (45)$$

and to determine ψ the equation is then

$$\nabla^2(\psi) = -c^2 \sin 2\eta. \quad (46)$$

By direct integration of this equation, or by (21) above, we find

$$\psi_E = \frac{c^2}{4} \sin 2\eta \quad (47)$$

and

$$\psi = \frac{c^2}{4} \sin 2\eta + A\eta. \quad (48)$$

If the boundary surface be represented by $\eta = \eta_0$, the condition that velocity is zero at $\eta = \eta_0$ gives

$$A = -\frac{c^2}{2} \cos 2\eta_0. \quad (49)$$

and the required solution is finally given by

$$\psi = \frac{c^2}{4} (\sin 2\eta - 2 \cos 2\eta_0 \cdot \eta) \quad (50)$$

and

$$\left. \begin{aligned} u &= \frac{c^2(\cos 2\eta - \cos 2\eta_0)}{2\sqrt{c^2(\cosh 2\xi - \cosh 2i\eta)/2}} \\ &= \frac{c(\cos^2 \eta - \cos^2 \eta_0)}{\sqrt{(\cosh^2 \xi - \cos^2 \eta)}} \end{aligned} \right\} \dots \dots \dots (51)$$

We obtain also the values

$$\zeta = \frac{\sin 2\eta}{\cosh^2 \xi - \cos^2 \eta}; \quad \dots \dots \dots (52)$$

$$\frac{p}{\mu} = \frac{\sinh 2\xi}{\cosh^2 \xi - \cos^2 \eta} \quad \dots \dots \dots (53)$$

When $\eta_0=0$, the flow takes place through an aperture of width $2c$ in an infinite plane sheet of negligible thickness. For any point in the aperture ($x < c$, $y=0$) we have

$$\cos \eta = \frac{x}{c}, \quad \cosh \xi = 1. \quad \dots \dots \dots (54)$$

The velocity distribution across the aperture is given by

$$u = - \frac{a^2 - x^2}{\sqrt{c^2 - x^2}} \quad \dots \dots \dots (55)$$

for an aperture of width $2a$, corresponding to the boundary $\eta = \eta_0$ where $\cos \eta_0 = \frac{a}{c}$. For an aperture of width $2c$ in the plane $y=0$, $\eta_0=0$, and $a=c$, so that

$$u = - \sqrt{c^2 - x^2}, \quad \dots \dots \dots (56)$$

a result given by Smoluchowski⁽³⁾.

Since ψ depends upon η only the streamlines in this case are all hyperbolas represented by (42) above for all values of η . At all points at exceedingly large distances from the aperture each hyperbola approaches closely to its asymptote, and the flow is then practically radial and directed towards the centre point of the aperture. Thus at large distances from the origin the two solutions, for converging plane boundaries and for hyperbolic boundaries respectively, come into agreement.

Flow into a Parallel Channel.

We now define the variables (ξ , η) by means of the equation

$$x + iy = \frac{h}{\pi} + \frac{h}{c}(\xi + i\eta) + \frac{h}{\pi}e^{\frac{\pi}{c}(\xi + i\eta)} = f(\xi + i\eta). \quad \dots \dots \dots (57)$$

This gives

$$x = \frac{h}{\pi} + \frac{h}{c}\xi + \frac{h}{\pi}e^{\frac{\pi}{c}\xi} \cos \frac{\pi}{c}\eta. \quad \dots \dots \dots (58)$$

$$y = \frac{h}{c}\eta + \frac{h}{\pi}e^{\frac{\pi}{c}\xi} \sin \frac{\pi}{c}\eta. \quad \dots \dots \dots (59)$$

When $\eta=0$; $y=0$. Thus the X-axis is locus $\eta=0$. When

$$\eta=\pm c; \quad y=\pm h, \quad x=\frac{h}{\pi}+\frac{h}{c}\xi-\frac{h}{\pi}e^{\frac{\pi}{c}\xi}.$$

In this let

$$\xi=+\infty, \quad x=-\infty,$$

$$\xi=0, \quad x=0,$$

$$\xi=-\infty, \quad x=-\infty.$$

The locus $\eta=0$ is midline of the system, and the locus $\eta=c$ consists of two lines, $y=\pm h$, parallel to it, each extending from $-\infty$ to 0 and from 0 back to $-\infty$, as ξ passes from $+\infty \rightarrow 0 \rightarrow -\infty$. Each locus may be regarded as a straight line doubled back on itself. The locus $\eta=\pm c'$, where $c'<c$, are streamlines for a perfect fluid flowing from an open space into a parallel channel. Some of these are shown in a diagram given by Lamb in 'Hydrodynamics,' Edition 6, p. 74, where this problem of flow of a perfect liquid is discussed. The diagram is here reproduced for convenience.

Putting $H^2=f'(\xi+i\eta)f'(\xi-i\eta), \dots \dots \dots (60)$

we have $H^2=\frac{h^2}{c^2}\left\{1+e^{\frac{\pi}{c}(\xi+i\eta)}\right\}\left\{1+e^{\frac{\pi}{c}(\xi-i\eta)}\right\}$

or $H^2=\frac{h^2}{c^2}\left\{1+2e^{\frac{\pi}{c}\xi}\cos\frac{\pi}{c}\eta+e^{\frac{2\pi}{c}\xi}\right\} \dots \dots \dots (61)$

In accordance with (23) above we now write

$$\zeta=4\left\{\frac{F'(\xi+i\eta)}{\frac{h}{c}\left(1+e^{\frac{\pi}{c}(\xi+i\eta)}\right)}-\frac{F'(\xi-i\eta)}{\frac{h}{c}\left(1+e^{\frac{\pi}{c}(\xi-i\eta)}\right)}\right\} \dots \dots \dots (62)$$

In this case the boundary of the fluid is represented by the curves $\eta=\pm c'$, where $c'<c$, and the problem practically reduces to that of determining the form of F , so that the condition of zero motion at each point of the boundary is fulfilled.

Now, considering any boundary curve $\eta=\pm c'$: at $\xi=-\infty$, the boundary walls form a parallel channel defined by the lines $y=\pm\frac{c'}{c}h$.

For any point within the channel where the walls are nearly parallel, x is large and negative. Thus in passing along any such curve $\xi=+\infty$ at points at infinite distance outside the channel, $\xi=-\infty$ at points at infinite distance from the origin within the channel, and $\xi=0$ at some point near, and to the right of the origin.

The $\xi=\text{constant}$, and $\eta=\text{constant}$, curves cut each other at right angles.

Assuming that a steady flow into the channel has been established, the flow within the channel, and at infinite distance from the origin, must

correspond to that in a straight channel of infinite length. Thus at $\xi = -\infty$, $y = \pm \frac{c'}{c}h$ represents the boundary walls for large negative values of x , and the solution required must approximate to that indicated by

$$u = \frac{\partial \psi}{\partial y} = \frac{u_0}{\alpha^2 h^2} (y^2 - \alpha^2 h^2), \text{ where } \alpha = \frac{c'}{c}, \quad (63)$$

and $-u_0$ is the velocity of flow along the mid-line of the channel. The approximate values of ψ and ζ corresponding to this solution are

$$\psi = \frac{u_0}{\alpha^2 h^2} \left(\frac{y^3}{3} - \alpha^2 h^2 y \right) \quad (64)$$

$$\text{and} \quad \zeta = -\frac{2u_0}{\alpha^2 h^2} y = -\frac{2u_0}{\alpha^2 h^2} \cdot \frac{h}{c} \eta = -\frac{2u_0 c}{c'^2 h} \eta \quad (65)$$

respectively.

Using this as a guide, we take as a possible form of F'

$$\frac{4F'(\xi + i\eta)}{\frac{h}{c} \left\{ 1 + e^{\frac{\pi}{c}(\xi + i\eta)} \right\}} = \frac{1}{i} \left\{ (\xi + i\eta) - A_1 e^{\frac{\pi}{c}(\xi + i\eta)} \right\}, \quad (66)$$

which gives

$$F(\xi + i\eta) = \frac{h}{4ci} \left[\frac{1}{2} w^2 + \frac{c}{\pi} w e^{\frac{\pi}{c} w} - \frac{c}{\pi} \left(\frac{c}{\pi} + A_1 \right) e^{\frac{\pi}{c} w} - A_1 \frac{c}{2\pi} e^{\frac{2\pi}{c} w} \right], \quad (67)$$

where $w = (\xi + i\eta)$. From (21) above we obtain the particular integral of (17) in the form

$$\psi_P = \frac{2}{i} \times \text{imaginary part of} \quad \left[\frac{h}{\pi} + \frac{h}{c} w + \frac{h}{\pi} e^{\frac{\pi}{c} w} \right] F(w). \quad (68)$$

In this expression the terms depending on $\frac{h}{\pi}$ are potential terms, and these may be regarded as included in the complementary function. With the omission of potential terms from (68) we may write the particular integral in the form

$$\begin{aligned} \psi_P = -\frac{h^2}{2c^2} & \left[\frac{\xi^2 \eta + \eta^3}{2} + \frac{c}{2\pi} (\xi^2 + 3\eta^2) e^\alpha \sin \beta + \frac{c}{\pi} \xi \eta e^\alpha \cos \beta \right. \\ & - \frac{c}{\pi} \left(\frac{c}{\pi} + A_1 \right) \{ \xi e^\alpha \sin \beta - \eta e^\alpha \cos \beta \} \\ & - A_1 \frac{c}{2\pi} e^{2\alpha} \{ \xi \sin 2\beta - \eta \cos 2\beta \} \\ & \left. + \frac{c^2}{\pi^2} \eta e^{2\alpha} - A_1 \frac{c^2}{2\pi^2} e^{2\alpha} \sin \beta \right], \quad (69) \end{aligned}$$

where

$$\alpha = \frac{\pi}{c} \xi, \quad \beta = \frac{\pi}{c} \eta.$$

By the addition of suitable potential terms the terms containing ξ and ξ^2 in the above can be removed. Any function of w , such as w^n , $w^n e^{\frac{\pi w}{c}}$,

yields two potential terms, one even with respect to η and one odd. In all cases of symmetrical flow it is the odd function which is required. For example, if we add the terms

$$\frac{h^2}{4c^2} \left(\xi^2 \eta - \frac{\eta^3}{3} \right) : \frac{h^2}{4\pi c} \{ (\xi^2 - \eta^2) e^\alpha \sin \beta + 2\xi \eta e^\alpha \cos \beta \} : \\ - \frac{h^2}{2\pi c} \left(\frac{c}{\pi} + A_1 \right) e^\alpha \{ \xi \sin \beta + \eta \cos \beta \} : - \frac{h^2 A_1}{2\pi} e^{2\alpha} \{ \xi \sin 2\beta + \eta \cos 2\beta \},$$

the expression representing the particular integral becomes

$$\psi_P = - \frac{h^2}{3c^2} \left\{ \eta^3 + 3 \frac{c}{\pi} \eta^2 e^\alpha \sin \beta + 3 \frac{c}{\pi} \left(\frac{c}{\pi} + A_1 \right) \eta e^\alpha \cos \beta \right. \\ \left. + \frac{3}{2} A_1 \frac{c}{\pi} \eta e^{2\alpha} \cos 2\beta + \frac{3}{2} \frac{c^2}{\pi^2} \eta e^{2\alpha} \right. \\ \left. - \frac{3}{4} A_1 \frac{c^2}{\pi^2} e^{3\alpha} \sin \beta \right\} (70)$$

The flow being symmetrical on the two sides of the line $\eta=0$, ψ must be an odd function of η . Hence we may take as complementary function ψ_c , corresponding to ψ_P above

$$\psi_c = - \frac{h^2}{3c^2} \{ A_0' \eta + A_1' e^\alpha \sin \beta + A_2' e^{2\alpha} \sin 2\beta + A_3' e^{3\alpha} \sin 3\beta \}, . . (71)$$

where A_0' , A_1' , etc. are constants undetermined, and the general integral of the equation for ψ is then

$$\psi = \psi_P + \psi_c, \text{ as given above.} (25)$$

As this expression satisfies the fundamental hydrodynamical equations it remains to examine whether it may also satisfy the boundary conditions at particular boundaries. At any boundary defined by $\eta=\text{constant}$, η_0 , the required conditions are

$$u = \frac{1}{H} \frac{\partial \psi}{\partial \eta} = 0 \text{ and } v = - \frac{1}{H} \frac{\partial \psi}{\partial \xi} = 0, \text{ at } \eta = \eta_0. . . . (72)$$

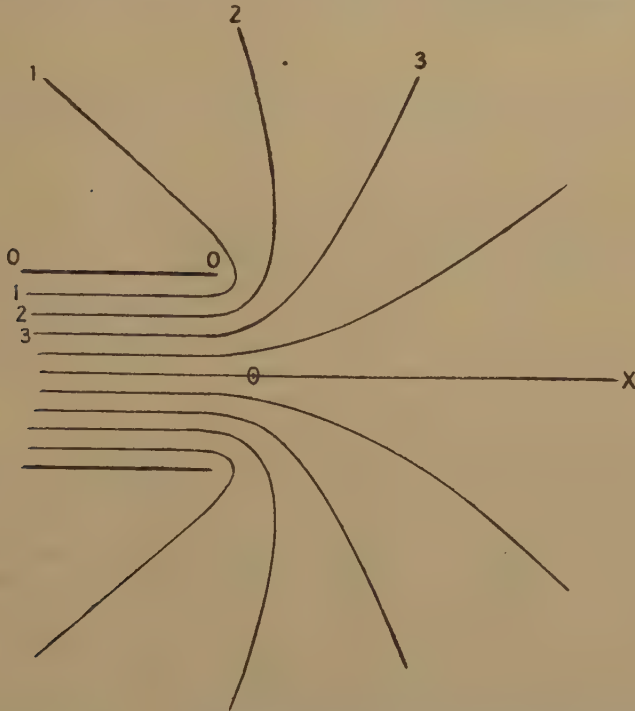
These conditions are satisfied by (70) and (71) for three separate boundaries, by proper choice of A_1 .

Boundary $\eta=c$. By applying these conditions at this boundary, we obtain $\left(A_1 = -\frac{c}{\pi} \right)$.

$$\psi = - \frac{h^2}{3c^2} \left\{ \eta^3 + 3 \frac{c}{\pi} \eta^2 e^\alpha \sin \beta - \frac{3c^2}{2\pi^2} \eta e^{2\alpha} \cos 2\beta + \frac{3c^2}{2\pi^2} \eta e^{2\alpha} + \frac{3c^3}{4\pi^3} e^{3\alpha} \sin \beta \right. \\ \left. - 3c^2 \eta - 3 \frac{c^3}{\pi} e^\alpha \sin \beta - \frac{3}{4} \frac{c^3}{\pi^3} e^{2\alpha} \sin 3\beta \right\} \\ = - \frac{h^2}{3c^2} \left\{ \left(\eta + \frac{c}{\pi} e^\alpha \sin \beta \right)^3 - 3c^2 \left(\eta + \frac{c}{\pi} e^\alpha \sin \beta \right) \right\} \\ = - \frac{c}{3h} \{ y^3 - 3h^2 y \}, (73)$$

which, with a multiplier, comes into agreement with (64). We have thus recovered the steady motion solution representing flow in a channel of infinite length as the only solution possible representing flow from the open into an open channel, having perfectly parallel walls as boundaries, in accordance with Helmholtz's theorem.

Boundary $\eta = \eta_0$, where $\tan 2\beta_0 = \frac{2\pi}{c} \eta_0$. Here $2\beta_0 = 4.493$ radians, and the boundary is represented in the accompanying figure by



a curve of the type 11. The values of the constants found by means of the boundary conditions are

$$\begin{aligned} A_1 &= 0; \quad A_0' = -3\eta_0^2; \quad A_1' = -3\frac{c^2}{\pi^2}\beta_0(\beta_0 + \cot \beta_0); \\ A_2' &= -\frac{3}{2}\frac{c^2}{\pi^2}\frac{\beta_0}{\sin 2\beta_0}, \quad \dots \dots \dots (74) \end{aligned}$$

and the solution is expressed by

$$\begin{aligned} \psi = -\frac{h^2}{3c^2} \left\{ \eta^3 + 3\frac{c}{\pi}\eta^2 e^\alpha \sin \beta + 3\frac{c^2}{\pi^2}\eta e^\alpha \cos \beta + \frac{3}{2}\eta^2 e^{2\alpha} \right. \\ \left. + A_0'\eta + A_1'e^\alpha \sin \beta + A_2'e^{2\alpha} \sin 2\beta \right\}. \quad \dots \quad (75) \end{aligned}$$

$$\xi = 2\eta. \quad \dots \dots \dots (76)$$

$$\frac{p}{\mu} = -2\xi \quad \dots \dots \dots (77)$$

Agreement with (64) and (65) at points well within the parallel channel is again obtained by means of a negative multiplier. Taking the boundary line $\eta = \eta_0$ or any of the curves $\eta = \text{constant}$, the pressure along it varies from $-\infty$ at points at infinite distance within the channel to zero near its entrance, and then to $+\infty$ at points at infinite distance outside the channel.

Boundary $\eta = \frac{c}{2}$. In this case the values of the constants appearing in (70) and (71) determined in accordance with the boundary conditions, are given by

$$A_1 = \frac{c}{\pi} : A_0' = -\frac{3}{4}c^2 : A_1' = -\frac{3}{4}\frac{c^2}{\pi} : A_2' = 0 : A_3' = -\frac{3}{4}\frac{c^3}{\pi^3}. \quad (78)$$

The solution is accordingly expressed by

$$\psi = -\frac{h^2}{3c^2} \left\{ \eta^3 + 3\frac{c}{\pi}\eta^2 e^\alpha \sin \beta + 6\frac{c^2}{\pi^2}\eta e^\alpha \cos \beta + \frac{3c^2}{2\pi^2}\eta e^{2\alpha}(1 + \cos \beta) - \frac{3}{4}\frac{c^3}{\pi^3}e^{3\alpha} \sin \beta - \frac{3}{4}c^2\eta - \frac{3}{4}\frac{c^3}{\pi}e^\alpha \sin \beta - \frac{3}{4}\frac{c^3}{\pi^3}e^{3\alpha} \sin 3\beta \right\}. \quad (79)$$

$$\zeta = 2 \left(\eta - \frac{c}{\pi} e^\alpha \sin \beta \right). \quad (80)$$

$$\frac{p}{\mu} = -2 \left(\xi - \frac{c}{\pi} e^\alpha \cos \beta \right). \quad (81)$$

In order to bring this solution into conformity with the solution (64) and (65) at points at infinite distance from the origin within the parallel channel the multiplier, $-\frac{u_0}{\frac{1}{4}h^2} \cdot \frac{h}{c}$, is required. This solution presents interesting features in the pressure distribution⁽⁴⁾, and possibly also with respect to flow, the investigation of which may be carried out later.

The above solutions are obtained from equation (69) above, and they correspond to the values of the constant A_1 given by $-\frac{c}{\pi}$, 0, $+\frac{c}{\pi}$ respectively. They refer to the boundary curves represented by the values of η_0 given by $\eta_0 = c$, η_0 defined by $\frac{2\pi}{c}\eta_0 = \tan \frac{2\pi}{c}\eta_0$, and $\eta_0 = \frac{c}{2}$ respectively. In the figure reproduced from Lamb's 'Hydrodynamics,' 3rd ed. p. 71, to indicate the curves $\eta = \text{constant}$, the first of the above curves is the two parallel boundaries 0 0, and the second corresponds closely to the curve marked 11. The third is the curve defined by equations

$$x = \frac{h}{\pi} + \frac{h}{c}\xi : y = \frac{h}{2} + \frac{h}{\pi}e^{\frac{\pi}{2}\xi}, \quad (82)$$

and is represented by a curve lying between 2 2 and 3 3,

A more comprehensive solution, referring to any boundary curve of the type $\eta = \eta_0$, ($\eta_0 < c$), can be obtained by assuming, instead of the right-hand side of (66) above, an expression of the form

$$\frac{1}{i} \left\{ w - A_1 e^{\frac{\pi}{c} w} - A_2 e^{\frac{2\pi}{c} w} - \dots, \text{etc.} \right\}, \quad \dots \quad (83)$$

in accordance with which ζ is expressed by an infinite series of terms. This expression defines the function F , as in (66) above; and each term having coefficient A_n contributes a term ψ_n to the particular integral of (17), where ψ_n is given by

$$\begin{aligned} \psi_n = & -A_n \frac{h^2}{c^2} e^{n\alpha} \left\{ \frac{c}{\pi} \eta \left[\frac{1}{n} \cos n\beta + \frac{1}{(n+1)} e^\alpha \cos (n+1)\beta \right] \right. \\ & \left. - \frac{c^2}{2\pi^2} \left[\frac{1}{n} e^\alpha \sin (n-1)\beta + \frac{1}{(n+1)} e^{2\alpha} \sin n\beta \right] \right\}. \quad \dots \quad (84) \end{aligned}$$

The contribution to the particular integral arising from the first two terms of (83), namely $\psi_0 + \psi_1$, is already given on the right-hand side of (70).

The particular integral is now expressed by

$$\psi_P = \sum_0^\infty \psi_n, \quad \dots \quad (85)$$

where the terms are defined by (84) and (70) above.

As the corresponding complementary function, consisting of potential terms, odd with respect to η , we may take

$$\psi_c = -\frac{h^2}{3c^2} \left[A_0' \eta + \sum_1^\infty A_n' e^{n\alpha} \sin n\beta \right]. \quad \dots \quad (86)$$

The complete solution is represented by

$$\psi = (\psi_P + \psi_c) \text{ as before } \quad \dots \quad (25)$$

and the constants are to be determined by means of the two boundary conditions. For convenience the terms in the series (25) representing the general integral may be arranged in ascending powers of e^α , thus

$$\begin{aligned} \psi = & -\frac{h^2}{3c^2} \left[(\eta^3 + A_0' \eta) + e^\alpha \left(\frac{3c}{\pi} \eta^2 \sin \beta + \frac{3c}{\pi} \left(\frac{c}{\pi} + A_1 \right) \eta \cos \beta + A_1' \sin \beta \right) \right. \\ & + \frac{3}{2} e^{2\alpha} \left(\frac{c^2}{\pi^2} \eta + (A_1 + A_2) \frac{c}{\pi} \eta \cos 2\beta + A_2' \sin 2\beta \right) \\ & + \frac{3}{2} \sum_3^\infty e^{n\alpha} \left\{ -\frac{(A_{n-2} + A_{n-1})}{n-1} \frac{c^2}{\pi^2} \sin (n-2)\beta \right. \\ & \left. \left. + \frac{2}{n} (A_{n-1} + A_n) \frac{c}{\pi} \eta \cos n\beta + A_n' \sin n\beta \right\} \right]. \quad \dots \quad (87) \end{aligned}$$

The conditions to be fulfilled at any boundary $\eta = \eta_0$ yield equations which suffice to determine the values of the constants, $A_0, A_0', \dots, A_n, A_n'$, corresponding to each index n appearing in each of the multipliers of

$e^{n\alpha}$ in the expression representing ψ . The values so obtained for the coefficients in (83) above are given by

$$A_1 = \frac{c \sin 2\beta_0 - 2\beta_0 \cos 2\beta_0}{\pi (2\beta_0 - \sin 2\beta_0)} \quad \dots \quad (88)$$

$$(A_1 + A_2) = \frac{2c \sin 2\beta_0 - 2\beta_0 \cos 2\beta_0}{\pi (4\beta_0 - \sin 4\beta_0)} \quad \dots \quad (89)$$

$$(A_2 + A_3) = (A_1 + A_2) \frac{c \frac{1}{2} \sin 4\beta_0 - \sin 2\beta_0}{\pi \frac{1}{3}(6\beta_0 - \sin 6\beta_0)} \quad \dots \quad (90)$$

$$A_n + A_{n+1} = (A_{n-1} + A_n) \frac{c \frac{1}{n} \sin 2n\beta_0 - \sin 2\beta_0}{\pi \frac{1}{n+1} (2(n+1)\beta_0 - \sin 2(n+1)\beta_0)} \quad \dots \quad (91)$$

Of the constants A'_0, A'_1 , etc. $A'_0 = -3\eta_0^2$; and the remaining coefficients are to be determined by equating to zero the expressions multiplying each $e^{n\alpha}$ in (87) above. Thus the solutions representing flow into a parallel channel within the boundaries specified by $\eta = \eta_0$ have been found; and the fluid movement represented is unique in accordance with the well-known theorem of Helmholtz⁽⁶⁾. The study of these solutions in detail must be deferred to a later opportunity.

References.

- (1) Sampson, Phil. Trans. 182 A (1891).
- (2) Lord Rayleigh, Phil. Mag. (Oct. 1893).
- (3) Smoluchowski, *Bull. de l'Acad. des Sc. de Cracovie* (Jan. 1907).
- (4) Blasius, *Zeit. der Math.* lviii. (1910).
- (5) Harrison, Proc. Camb. Phil. Soc. xix. (1916-19).
- (6) Helmholtz, *Abh. t. 1* (1868).
- (7) 'Modern Fluid Dynamics.'

XXXII. *Coefficients for Numerical Integration with Central Differences.*

By HERBERT E. SALZER * †.

[Received August 25, 1943.]

THE following table lists the coefficients in the formula for numerical integration employing central differences through the fifty-fifth order. Previous calculations have gone only as far as coefficients of the seventh central difference. The coefficients M_{2s} occur in the following formula :

$$\frac{1}{h} \int_a^{a+nh} f(x) dx = \left[\frac{1}{2} f(a) + f(a+h) + \dots + f(a+(n-1)h) + \frac{1}{2} f(a+nh) \right] \\ + \sum_{s=1}^{m-1} M_{2s} [\mu \delta^{2s-1} f(a+nh) - \mu \delta^{2s-1} f(a)] + nh^{2m} M_{2m} f^{(2m)}(\xi). \quad (1)$$

* Communicated by the Author.

† Mathematical Tables Project, National Bureau of Standards, U.S.A.

The coefficient M_{2s} is identical with $B_{2s}^{(2s)}(s)/(2s)!$ where $B_{2s}^{(2s)}(s)$ is the $(2s)$ th Bernoulli polynomial of order $2s$ for argument equal to s . Equation (1) is known as the Gauss-Encke formula and also as the second Gaussian summation-formula. The following properties of these coefficients, given in (2)-(7), are all demonstrated in Steffensen :

[illegible]

$$M_{2s} = \frac{2}{(2s)!} \sum_{r=0}^{r=s} \frac{D^{2r+1} O_{[2s+1]}}{2^{2r+1} (2r+1) (2r+1)!}, \quad \dots \quad (3)$$

where $D^{2r+1}O^{[2s+1]}$ is the $(2r+1)$ th derivative of the central factorial polynomial

$$x^{[2s+1]} \equiv x(x^2 - 1/4)(x^2 - 9/4) \dots \left[x^2 - \frac{(2s-1)^2}{4} \right]$$

for $x=0$.

$$M_{2s} = \frac{2}{(2s)!} \int_0^1 \frac{\Gamma(x+s+\frac{1}{2})}{\Gamma(x-s+\frac{1}{2})} dx, \quad . \quad . \quad . \quad . \quad . \quad (4)$$

which is identical with Milne-Thomson's

$$M_{2s} = \frac{1}{(2s)!} \int_0^1 (x+s-1)(x+s-2) \dots (x-s) dx, \quad . \quad . \quad . \quad (4a)$$

$$|M_{2s}| \leq \binom{2s}{s} 2^{-4s}. \quad \dots \dots \dots (5)$$

For large s ,

$$M_{2s} \sim \frac{(-1)^s}{2^{2s-1} \pi^{3/2}} \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (6)$$

(At $s=10$, this formula is accurate to about 1 per cent.).

$$M_{2s} = \frac{1^2}{3! 2^2} M_{2s-2} - \frac{(1.3)^2}{5! 2^4} M_{2s-4} + \frac{(1.3.5)^2}{7! 2^6} M_{2s-6} - \dots$$

$$+ \frac{(-1)^s [1.3 \dots (2s-3)]^2}{(2s-1)! 2^{2s-2}} M_2 + \frac{(-1)^s [1.3 \dots (2s-1)]^2}{(2s-1)! (2s+1) 2^{2s}}. \quad (7)$$

The first twenty coefficients given below were checked, using their exact values, in the cumulative recursion formula (7). All values were checked both by differencing the ratios M_{2s}/M_{2s+2} and by actual integrations.

Table of Coefficients.

$$M_2 - \frac{1}{12}.$$

$$M_4 + \frac{11}{720}.$$

$$M_6 = -\frac{191}{60480}.$$

M_8	$+\frac{2497}{3628800}.$
M_{10}	$-\frac{14797}{95800320}.$
M_{12}	$+\frac{924 \quad 27157}{261 \quad 53487 \quad 36000}.$
M_{14}	$-\frac{367 \quad 40617}{448 \quad 34549 \quad 76000}.$
M_{16}	$+\frac{6 \quad 14309 \quad 43169}{32 \quad 01186 \quad 85286 \quad 40000}.$
M_{18}	$-\frac{2313 \quad 39458 \quad 92303}{51090 \quad 94217 \quad 17094 \quad 40000}.$
M_{20}	$+\frac{1639 \quad 96886 \quad 81447}{152579 \quad 28431 \quad 37024 \quad 00000}.$
M_{22}	$-\cdot 00000 \quad 00256 \quad 38286 \quad 986$
M_{24}	$\cdot 00000 \quad 00061 \quad 40295 \quad 342$
M_{26}	$-\cdot 00000 \quad 00014 \quad 75583 \quad 052$
M_{28}	$\cdot 00000 \quad 00003 \quad 55628 \quad 077$
M_{30}	$-\cdot 00000 \quad 00000 \quad 85924 \quad 015$
M_{32}	$\cdot 00000 \quad 00000 \quad 20805 \quad 605$
M_{34}	$-\cdot 00000 \quad 00000 \quad 05047 \quad 538$
M_{36}	$\cdot 00000 \quad 00000 \quad 01226 \quad 642$
M_{38}	$-\cdot 00000 \quad 00000 \quad 00298 \quad 549$
M_{40}	$\cdot 00000 \quad 00000 \quad 00072 \quad 762$
M_{42}	$-\cdot 00000 \quad 00000 \quad 00017 \quad 755$
M_{44}	$\cdot 00000 \quad 00000 \quad 00004 \quad 338$
M_{46}	$-\cdot 00000 \quad 00000 \quad 00001 \quad 061$
M_{48}	$\cdot 00000 \quad 00000 \quad 00000 \quad 260$
M_{50}	$-\cdot 00000 \quad 00000 \quad 00000 \quad 064$
M_{52}	$\cdot 00000 \quad 00000 \quad 00000 \quad 016$
M_{54}	$-\cdot 00000 \quad 00000 \quad 00000 \quad 004$
M_{56}	$\cdot 00000 \quad 00000 \quad 00000 \quad 001$

References.

- D. K. Sen, "Interpolation and Summation Formulas and the Properties of Factorials," Journ. Univ. Bombay, vol. xi. pt. 5, pp. 5-23 (Mar. 1943).
 J. F. Steffensen, 'Interpolation,' pp. 109-110, 191 (1927).
 L. M. Milne-Thomson, 'Calculus of Finite Differences,' pp. 184-187 (1933).
 J. B. Scarborough, 'Numerical Mathematical Analysis,' pp. 124-131, 165-167 (1930).
 E. T. Whittaker, G. Robinson, 'Calculus of Observations,' pp. 146-149 (1937).
 Bulletin National Research Council, No. 92, p. 28 (1933). (Obvious misprint in M_8 .)

XXXIII. *Equilibrium of a Thin Slab on an Elastic Foundation of Finite Depth.*

By A. H. A. HOGG *.

[Received August 25, 1943.]

1. *Introduction.*

THE stresses in a loaded slab resting on compressible ground are of interest in connection with the design of road slabs and raft footings. A certain amount of experimental work⁽¹⁾ on the subject has been published, but the interpretation of the results has suffered owing to the lack of an adequate body of theory for comparison.

2. *Previous Work.*

The principal theoretical investigations of this subject are :

(i.) " Westergaard's Theory," in which the deflection is assumed to be proportional to the intensity of load at the point considered⁽²⁾.

(ii.) A note by the present writer on the deflection of a thin slab of infinite extent resting on an elastic " half-space "⁽³⁾ ".

(iii.) A paper giving the solution for the two dimensional case corresponding to (ii.)⁽⁴⁾.

It is clear that the assumptions on which these theories are based differ in many ways from the conditions which occur in practice. So far, no detailed discussion of the experimental work has been published ; but a limited examination suggests that it would be advantageous to have available the theoretical results for the case considered in the present paper.

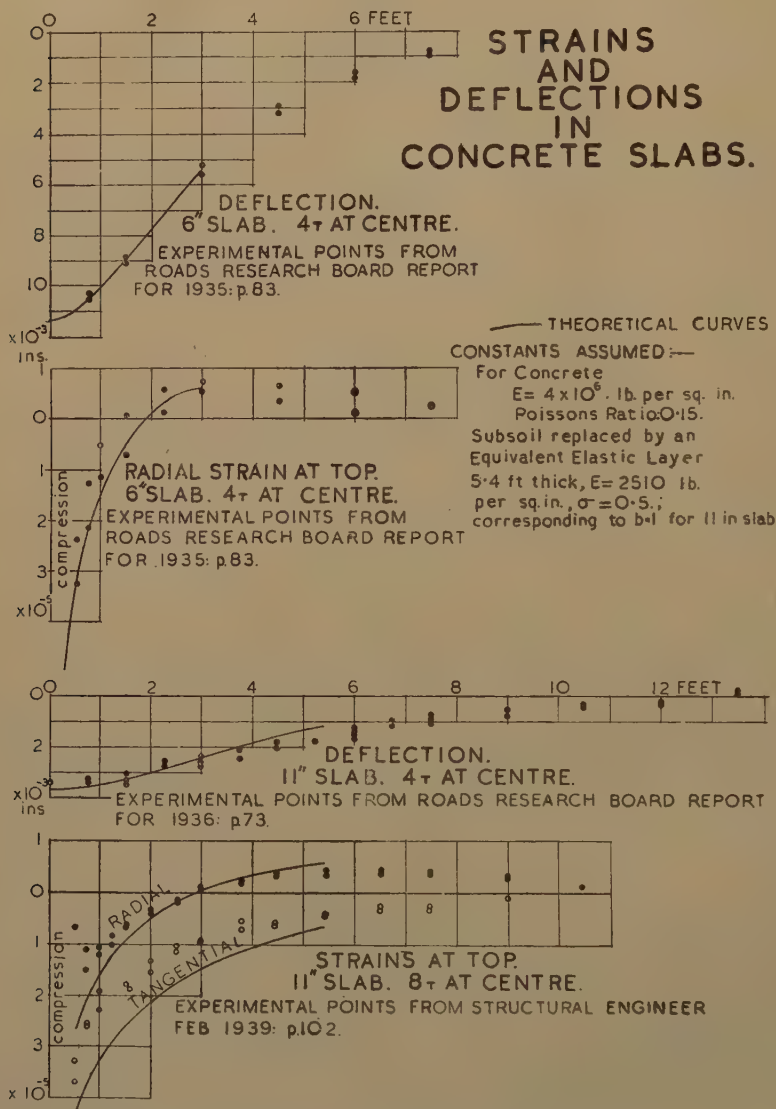
3. *Subject of Paper.*

The system considered is composed of a uniform layer of elastic material, resting on a perfectly rough rigid horizontal base. A thin elastic slab, of infinite extent, rests on the upper surface, and there is no slip between the slab and the elastic material. The slab carries a single concentrated load. System (ii.) above is a special case of this. The exact mathematical solution for the more general case is obtained in the form of an infinite integral. Approximate values have been calculated by numerical integration for the range which is likely to be of use in practice.

* Communicated by the Author,

No attempt is made to compare the experimental and theoretical results in detail. It seems better to leave this for separate discussion, since the question as to how far the ideal assumptions made apply to a practical case is a problem of an entirely different character from that

Fig. 1.



considered in this paper. But in order to illustrate the sort of agreement which may be expected, the published results of experiments made by the Building Research Station are shown, together with appropriate curves based on this theory (fig. 1). In this particular case the "equivalent elastic layer" seems unlikely to bear any close relationship to the

actual foundation. It forms, however, a method of representing the relationship between load and deflection on the surface of the subgrade.

4. Theory.

In the analysis which follows the system has circular symmetry about the axis $r=0$. Z is taken as positive downwards.

The notation of Love's 'Elasticity' is used for the elastic constants, displacements, and stresses in the elastic layer and the suffix o indicates reference to $z=0$.

In addition :

h =depth of elastic layer,

$R = \frac{2E(1-\sigma)}{(1+\sigma)(3-4\sigma)}$ an elastic constant,

depending on the properties of the elastic layer.

D =Flexural rigidity of slab.

$b^3 = Rr^3/D$.

Π =total load on slab.

p =load on slab per unit area.

For the case of circular symmetry, any distribution of stress through an elastic medium can be expressed in the form ⁽⁵⁾

$$\widehat{zz} = - \int_0^\infty \{ [P_1 + R_1 kz - (1-2\sigma)S_1] \sinh kz + \{Q_1 - (1-2\sigma)R_1 + S_1 kz\} \cosh kz \} k^3 J_0(kr) dk, \quad (4.1)$$

$$U = \frac{1+\sigma}{E} \int_0^\infty \{ [P_1 + R_1 kz + S_1] \sinh kz + \{Q_1 + R_1 + S_1 kz\} \cosh kz \} k^2 J_1(kr) dk, \quad (4.2)$$

$$w = - \frac{1+\sigma}{E} \int_0^\infty \{ [Q_1 - 2(1-2\sigma)R_1 + S_1 kz] \sinh kz + \{P_1 + R_1 kz - 2(1-2\sigma)S_1\} \cosh kz \} k^2 J_0(kr) dk, \quad (4.3)$$

where P_1, Q_1, R_1, S_1 are constants.

The boundary conditions assumed are :—

At the bottom of the elastic layer ($z=h$), $U=0$ and $w=0$.

At the junction of the slab and foundation, ($z=0$), $U=0$ and $w=w_0$, and w_0 must have the same value for the slab and for the surface of the elastic layer.

The slab has flexural rigidity D , and carries a normal load

$$p = \int_0^\infty F(k) J_0(kr) dk.$$

The downward deflection of the slab is w_0 , and the foundation exerts a downward pull \widehat{zz}_0 on the slab.

Hence, from the usual expression for the displacement of a thin plate

$$\frac{1}{r} \frac{d}{dr} \left[r \frac{d}{dr} \left\{ \frac{1}{r} \frac{d}{dr} \left(r \frac{dw_0}{dr} \right) \right\} \right] = \frac{p + \widehat{z}_0}{D} \quad (4.4)$$

We require

$$w_0 = -\frac{1+\sigma}{E} \int_0^\infty \{P_1 - 2(1-2\sigma)S_1\} k^2 J_0(kr) dk.$$

Inserting the boundary conditions we have

$$\{P_1 + R_1 kh + S_1\} \sinh kh + \{Q_1 + R_1 + S_1 kh\} \cosh kh = 0, \quad (4.5)$$

$$\{Q_1 - 2(1-2\sigma)R_1 + S_1 kh\} \sinh kh + \{P_1 + R_1 kh - 2(1-2\sigma)S_1\} \cosh kh = 0, \quad (4.6)$$

$$Q_1 + R_1 = 0, \quad (4.7)$$

$$-\frac{1+\sigma}{E} \{P_1 - 2(1-2\sigma)S_1\} k^6 = \frac{1}{D} [F(k) - \{Q_1 - (1-2\sigma)R_1\} k^3], \quad (4.8)$$

whence

$$w_0 = + \frac{1}{D} \int_0^\infty \frac{F(k) k^2 J_0(kr) dk}{k^6 + \frac{2E(1-\sigma)}{D(1+\sigma)} k^3 \frac{(3-4\sigma) \sinh kh \cosh kh + kh}{(3-4\sigma)^2 \sinh^2 kh - k^2 h^2}}.$$

Writing m for kh and ρ for r/h .

$$w_0 = \frac{h^3}{D} \int_0^\infty \frac{F(k) J_0(m\rho) dm}{m^4 + \frac{2E(1-\sigma)}{D(1+\sigma)} m \frac{(3-4\sigma) \sinh m \cosh m + m}{(3-4\sigma)^2 \sinh^2 m - m^2}}.$$

5. Loading.

For a point load Π ⁽⁶⁾

$$F(k) = \frac{\Pi}{2\pi} k = \frac{\Pi}{2\pi} \frac{m}{h}, \quad (5.1)$$

and for a distributed load Π on a circular area radius a

$$F(k) = \frac{\Pi}{a\pi} J_1(ka) = \frac{\Pi}{a\pi} J_1\left(\frac{ma}{h}\right). \quad (5.2)$$

If we write

$$R = \frac{2E(1-\sigma)}{(1+\sigma)(3-4\sigma)}$$

and

$$\Phi(m) = \frac{\sinh m \cosh m + \frac{m}{3-4\sigma}}{\sinh^2 m - \frac{m^2}{(3-4\sigma)^2}} \quad (5.3)$$

we get for a point load

$$w_0 = \frac{\Pi h^2}{2\pi D} \int_0^\infty \frac{J_0(m\rho) dm}{m^3 + \frac{R h^3}{D} \Phi(m)}$$

For the special case $h \rightarrow \infty$ we already have the solution ⁽¹⁾

$$w_0 = \frac{\Pi}{(DR^2)^{\frac{1}{2}}} \cdot \frac{b^2}{2\pi} \int_0^\infty \frac{J_0(y) dy}{y^3 + b^3}, \quad \text{where } b^3 = \frac{R}{D} r^3,$$

and it will therefore be convenient to express the more general solution in the form

$$w_0 = \frac{\Pi}{(DR^2)^{\frac{1}{2}}} \frac{b^2}{2\pi\rho^2} \int_0^\infty \frac{J_0(m\rho) dm}{m^3 + \frac{Rh^3}{D} \Phi(m)}.$$

Differentiating with respect to r we have

$$\frac{dJ_0(m\rho)}{dr} = \frac{dJ_0(kr)}{dr} = -kJ_1(kr) = -\sqrt[3]{\frac{R}{D}} \frac{m\rho}{b} J_1(m\rho)$$

and
$$\frac{d^2J_0(m\rho)}{dr^2} = \frac{d^2J_0(kr)}{dr^2} = \frac{k}{r} \{J_1(kr) - krJ_0(kr)\}$$

$$= \left(\frac{R}{D}\right)^{\frac{2}{3}} \frac{m\rho}{b^2} \{J_1(m\rho) - m\rho J_0(m\rho)\}.$$

So
$$\frac{1}{r} \frac{dw_0}{dr} = -\frac{\Pi}{D} \frac{1}{2\pi\rho} \int_0^\infty \frac{mJ_1(m\rho) dm}{m^3 + \left(\frac{b}{\rho}\right)^3 \Phi(m)},$$

$$\frac{d^2w_0}{dr^2} = \frac{\Pi}{D} \frac{1}{2\pi\rho} \int_0^\infty \frac{\{mJ_1(m\rho) - m^2\rho J_0(m\rho)\} dm}{m^3 + \left(\frac{b}{\rho}\right)^3 \Phi(m)}.$$

These values are those produced by a concentrated load; if the load is uniformly distributed over a circle of radius a we must replace $\frac{\Pi}{2\pi} \frac{m}{h}$ by $\frac{\Pi}{a\pi} J_1\left(\frac{ma}{h}\right)$. Under the centre of the circle therefore we have

$$\frac{d^2w_0}{dr^2} = -\frac{\Pi}{D} \frac{h}{a\pi} \int_0^\infty \frac{mJ_1\left(\frac{ma}{h}\right) dm}{m^3 + \frac{Rh^3}{D} \Phi(m)}.$$

Hence it follows that $\frac{d^2w_0}{dr^2}$ under the centre of a load uniformly distributed over a circle of radius a is equal to *twice** the value of $\frac{1}{r} \frac{dw_0}{dr}$ produced at $r=a$ by an equal total load concentrated at $r=0$.

The exact integration of these results would seem to be a matter of some difficulty, but approximate values for

$$\frac{b^2}{2\pi\rho^2} \int_0^\infty \frac{J_0(m\rho) dm}{m^3 + \left(\frac{b}{\rho}\right)^3 \Phi(m)} = \frac{(DR^2)^{\frac{1}{2}}}{\Pi} w_0,$$

$$\frac{1}{2\pi\rho} \int_0^\infty \frac{m J_1(m\rho) dm}{m^3 + \left(\frac{b}{\rho}\right)^3 \Phi(m)} = -\frac{D}{\Pi} \frac{1}{r} \frac{dw_0}{dr},$$

$$\frac{1}{2\pi\rho} \int_0^\infty \frac{\{mJ_1(m\rho) - m^2\rho J_0(m\rho)\} dm}{m^3 + \left(\frac{b}{\rho}\right)^3 \Phi(m)} = \frac{D}{\Pi} \frac{d^2w_0}{dr^2},$$

over a limited range have been calculated by numerical integration.

* "Twice" is mistakenly omitted in the writer's previous paper (3), p. 579.

$$\frac{1}{2\pi\rho^2} \int_0^\infty \frac{\{m\rho^2 J_1(m\rho) - m^2\rho^2 J_0(m\rho)\} dm}{m^3 + \frac{b^3}{\rho^3} \Phi(m)}$$

	b	0.0	0.1	0.2	0.3	0.4
$\sigma = 0.5$	b/ρ	$\frac{1}{4}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	$\frac{1}{2}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	1	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	2	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	3	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	4	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	5	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	6	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
$\sigma = 0.4$	$\frac{1}{4}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	$\frac{1}{2}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	1	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	2	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	3	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	4	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	5	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	6	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
$\sigma = 0.3$	$\frac{1}{4}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	$\frac{1}{2}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	1	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	2	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	3	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	4	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	5	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	6	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
$\sigma = 0.2$	$\frac{1}{4}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	$\frac{1}{2}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	1	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	2	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	3	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	4	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	5	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	6	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
$\sigma = 0.1$	$\frac{1}{4}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	$\frac{1}{2}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	1	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	2	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	3	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	4	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	5	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	6	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
$\sigma = 0.0$	$\frac{1}{4}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	$\frac{1}{2}$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	1	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	2	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	3	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	4	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	5	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
	6	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$

$$\frac{1}{2\pi\rho} \int_0^\infty \frac{mJ_1(m\rho)dm}{m^3 + \left(\frac{b}{\rho}\right)^3} \Phi(\pi)$$

	b	0.0	0.1	0.2	0.3	0.4	
$\sigma=0.5$	b/ρ	$\frac{1}{2}$	$+\infty$.1517	.0975	.0691	.0471
	$\frac{1}{2}$	$+\infty$.1799	.1263	.0952	.0740	
	1	$+\infty$.2005	.1458	.1142	.0924	
	2	$+\infty$.2188	.1639	.1322	.1102	
	3	$+\infty$.2256	.1708	.1390	.1169	
	4	$+\infty$.2287	.1738	.1420	.1199	
	5	$+\infty$.2302	.1753	.1435	.1214	
	6	$+\infty$.2310	.1761	.1443	.1222	
$\star 7$	$+\infty$.2324	.1775	.1457	.1236		
		$-0.35 (\rho/b)^3$	$-0.35 (\rho/b)^3$	$-0.35 (\rho/b)^3$	$-0.35 (\rho/b)^3$		
$\sigma=0.4$	$\frac{1}{2}$	$+\infty$.1793	.1249	.0939	.0728	
	$\frac{1}{2}$	$+\infty$.1982	.1435	.1121	.0905	
	1	$+\infty$.2139	.1591	.1274	.1055	
	2	$+\infty$.2235	.1686	.1368	.1148	
	3	$+\infty$.2280	.1731	.1413	.1192	
	4	$+\infty$.2301	.1752	.1434	.1213	
	5	$+\infty$.2310	.1761	.1443	.1222	
	6	$+\infty$.2315	.1766	.1448	.1227	
$\star 7$	$+\infty$.2324	.1775	.1457	.1236		
		$-0.21 (\rho/b)^3$	$-0.21 (\rho/b)^3$	$-0.21 (\rho/b)^3$	$-0.21 (\rho/b)^3$		
$\sigma=0.3$	$\frac{1}{2}$	$+\infty$.1872	.1326	.1014	.0801	
	$\frac{1}{2}$	$+\infty$.2037	.1489	.1174	.0957	
	1	$+\infty$.2152	.1603	.1286	.1067	
	2	$+\infty$.2254	.1706	.1388	.1167	
	3	$+\infty$.2291	.1742	.1424	.1203	
	4	$+\infty$.2306	.1757	.1439	.1218	
	5	$+\infty$.2314	.1765	.1447	.1226	
	6	$+\infty$.2318	.1769	.1451	.1230	
$\star 7$	$+\infty$.2324	.1775	.1457	.1236		
		$-0.16 (\rho/b)^3$	$-0.16 (\rho/b)^3$	$-0.16 (\rho/b)^3$	$-0.16 (\rho/b)^3$		

$$\frac{b^2}{2\pi\rho^2} \int_0^\infty \frac{J_0(m\rho) \rho}{m^3 + \left(\frac{b}{\rho}\right)^3} dm$$

	$b \ 0.0$	0.1	0.2	0.3	0.4	0.5	
$\sigma=0.5$	$b/\rho \ \frac{1}{2}$.0088	.0079	.0062	.0042	.0023	.0003
	$\frac{1}{4}$.0247	.0236	.0215	.0188	.0159	.0129
	1	.0458	.0446	.0421	.0390	.0354	.0316
	2	.0813	.0800	.0773	.0737	.0695	.0648
	3	.1051	.1037	.1009	.0972	.0927	.0877
	4	.1218	.1204	.1176	.1138	.1092	.1041
	5	.1338	.1324	.1295	.1257	.1211	.1159
	6	.1428	.1414	.1385	.1347	.1300	.1248
	$\star 7$.1925	.1911	.1882	.1843	.1796	.1744
		$-0.311 \ \rho/b$	$-0.311 \ \rho/b$	$-0.311 \ \rho/b$	$-0.311 \ \rho/b$	$-0.311 \ \rho/b$	$-0.311 \ \rho/b$
				$+0.007 \ (\rho/b)^3$	$+0.016 \ (\rho/b)^3$	$+0.028 \ (\rho/b)^3$	$+0.043 \ (\rho/b)^3$
$\sigma=0.4$	$b/\rho \ \frac{1}{2}$.0310	.0299	.0278	.0252	.0223	.0193
	$\frac{1}{4}$.0469	.0457	.0433	.0402	.0367	.0329
	1	.0706	.0693	.0667	.0632	.0592	.0547
	2	.1038	.1024	.0997	.0960	.0916	.0868
	3	.1242	.1228	.1200	.1162	.1116	.1065
	4	.1379	.1365	.1336	.1298	.1252	.1200
	5	.1474	.1460	.1431	.1393	.1346	.1294
	6	.1543	.1529	.1500	.1461	.1415	.1363
	$\star 7$.1925	.1911	.1882	.1843	.1796	.1744
		$-0.236 \ \rho/b$	$-0.236 \ \rho/b$	$-0.236 \ \rho/b$	$-0.236 \ \rho/b$	$-0.236 \ \rho/b$	$-0.236 \ \rho/b$
				$+0.004 \ (\rho/b)^3$	$+0.010 \ (\rho/b)^3$	$+0.017 \ (\rho/b)^3$	$+0.027 \ (\rho/b)^3$
$\sigma=0.3$	$\frac{1}{2}$.0392	.0381	.0358	.0329	.0298	.0266
	$\frac{1}{4}$.0564	.0552	.0527	.0495	.0458	.0417
	1	.0821	.0808	.0781	.0747	.0706	.0661
	2	.1114	.1102	.1072	.1034	.0991	.0944
	3	.1333	.1319	.1291	.1253	.1207	.1155
	4	.1455	.1441	.1412	.1374	.1327	.1275
	5	.1538	.1524	.1495	.1456	.1409	.1357
	6	.1598	.1584	.1555	.1516	.1469	.1417
	$\star 7$.1925	.1911	.1882	.1843	.1796	.1744
		$-0.201 \ \rho/b$	$-0.201 \ \rho/b$	$-0.201 \ \rho/b$	$-0.201 \ \rho/b$	$-0.201 \ \rho/b$	$-0.201 \ \rho/b$
				$+0.003 \ (\rho/b)^3$	$+0.007 \ (\rho/b)^3$	$+0.012 \ (\rho/b)^3$	$+0.020 \ (\rho/b)^3$

$$\frac{R^2)^{\dagger}}{I} w_0.$$

0.6	0.7	0.8	0.9	1.0	2.0
011	-.0023	-.0032	-.0038	-.0040	—
100	-.0073	-.0048	-.0026	-.0007	—
278	-.0240	-.0204	-.0170	-.0139	—
601	-.0553	-.0505	-.0457	-.0411	-.0068
827	-.0775	-.0721	-.0668	-.0616	-.0196
989	-.0934	-.0879	-.0823	-.0768	-.0310
106	-.1050	-.0994	-.0937	-.0881	-.0405
194	-.1138	-.1081	-.1023	-.0967	-.0478
689	-.1632	-.1574	-.1515	-.1457	-.0948
11 ρ/b	-.0.311 ρ/b	-.0.311 ρ/b	-.0.311 ρ/b	-.0.311 ρ/b	-.0.311 ρ/b
62 $(\rho/b)^3$	+0.085 $(\rho/b)^3$	+0.111 $(\rho/b)^3$	+0.141 $(\rho/b)^3$	+0.173 $(\rho/b)^3$	+0.694 $(\rho/b)^3$
165	-.0139	-.0115	-.0093	-.0075	—
293	-.0256	-.0221	-.0188	-.0157	—
502	-.0457	-.0411	-.0367	-.0324	—
818	-.0766	-.0714	-.0663	-.0613	-.0216
014	-.0960	-.0905	-.0850	-.0795	-.0344
147	-.1092	-.1035	-.0978	-.0923	-.0452
240	-.1184	-.1127	-.1070	-.1013	-.0525
309	-.1252	-.1195	-.1137	-.1079	-.0583
689	-.1632	-.1574	-.1515	-.1457	-.0948
36 ρ/b	-.0.236 ρ/b	-.0.236 (ρ/b)	-.0.236 ρ/b	-.0.236 ρ/b	-.0.236 ρ/b
39 $(\rho/b)^3$	+0.053 $(\rho/b)^3$	+0.069 $(\rho/b)^3$	+0.089 $(\rho/b)^3$	+0.107 $(\rho/b)^3$	+0.428 $(\rho/b)^3$
234	-.0204	-.0175	-.0148	-.0125	—
378	-.0339	-.0300	-.0263	-.0228	—
615	-.0569	-.0523	-.0478	-.0435	-.0127
920	-.0868	-.0815	-.0762	-.0710	-.0289
103	-.1048	-.0993	-.0937	-.0881	-.0417
222	-.1165	-.1110	-.1052	-.0996	-.0511
304	-.1247	-.1190	-.1132	-.1075	-.0581
363	-.1306	-.1249	-.1191	-.1133	-.0634
689	-.1632	-.1574	-.1515	-.1457	-.0948
201 ρ/b	-.0.201 ρ/b	-.0.201 ρ/b	-.0.201 ρ/b	-.0.201 ρ/b	-.0.201 ρ/b
28 $(\rho/b)^3$	+0.038 $(\rho/b)^3$	+0.050 $(\rho/b)^3$	+0.063 $(\rho/b)^3$	+0.078 $(\rho/b)^3$	+0.313 $(\rho/b)^3$

Summary.

An expression is obtained for the displacement of a thin slab resting on an elastic layer of finite depth and carrying a single concentrated load. The results are expressed numerically for a range which may be of practical use.

References.

- (1) F. N. Sparkes, "Stresses in Concrete Road Slabs," 'Structural Engineer,' pp. 98-116 (Feb. 1939).
L. W. Teller, "Structural Design of Concrete Pavements."—Part III., 'Public Roads,' xvi. no. 10, pp. 201-221 (Dec. 1935).
- (2) H. M. Westergaard, 'Public Roads,' vii. p. 31 (1926-7).
- (3) A. H. A. Hogg, Phil. Mag. xxv. p. 576 (March 1938).
- (4) G. Bosson, Phil. Mag. xxvii. p. 37 (Jan. 1939).
- (5) W. Passer, "Druckverteilung durch eine Elastische Schicht," Akad der Wissenschaften in Wien. Math.-Natur. Klasse Abt. II a, 144 Band, 5 and 6 Heft. Wien, 1935.
- (6) K. Terazawa, "On the Elastic Equilibrium of a Semi-Infinite Solid," Journ. Coll. Science Imp. Univ. of Tokyo, xxxvii. Art. 7 (Dec. 1916).

XXXIV. *On Under-expansion in a Steam-nozzle, with Reference to an Unusual Case.*

By L. J. KASTNER, M.A.*

[Received August 27, 1943.]

WHEN a gaseous fluid, after expansion in a nozzle, issues from the nozzle with a pressure greater than the external pressure, it is said to be under-expanded, and its kinetic energy per unit mass is less than the value it would have taken up if the expansion had been complete. Assuming that the pressure ratio between the discharge and admission ends of the nozzle is less than the "critical" value, the nozzle should properly have a convergent-divergent form, and the loss of energy due to under-expansion, per unit mass of fluid passing, is dependent on the ratio between the actual nozzle exit area, and the exit area appropriate to complete expansion. In the limiting case, the exit area of the under-expanding nozzle will equal the area at the nozzle throat. This will correspond to the greatest degree of under-expansion, and the smallest area ratio possible, for any given pressure drop.

Consider the case of a convergent-divergent nozzle discharging into a region where the pressure is lower than that at the exit cross-section of the nozzle. Let the fluid in the nozzle expand according to the law $PV^n = \text{constant}$, so that the critical pressure P_c , which occurs at the narrowest cross-section, is given by the well-known expression

$$P_c = \left(\frac{2}{n+1} \right)^{\frac{n}{n-1}} P_1,$$

* Communicated by the Author.

where P_1 is the admission pressure. Let the pressure in the region into which the nozzle discharges be aP_c , the pressure at the exit cross-section of the nozzle being bP_c .

For a frictionless expansion in a correctly designed nozzle from a region where the pressure and specific volume are P_1, V_1 respectively to a region where the pressure is P_2 , the velocity at the exit cross-section will be

$$u = \sqrt{2g \cdot \frac{n}{n-1} \cdot P_1 V_1 \left[1 - \left(\frac{P_2}{P_1} \right)^{\frac{n-1}{n}} \right]}.$$

Hence, in the case under consideration, the maximum velocity which could be attained, if there were no under-expansion, equals

$$\begin{aligned} u_a &= \sqrt{2g \cdot \frac{n}{n-1} P_1 V_1 \left(1 - \left(a \left[\frac{2}{n+1} \right]^{\frac{n}{n-1}} \right)^{\frac{n-1}{n}} \right)} \\ &= \sqrt{2g \cdot \frac{n}{n-1} \cdot P_1 V_1 \left(1 - a^{\frac{n-1}{n}} \cdot \frac{2}{n+1} \right)}. \end{aligned}$$

However, when there is under-expansion, since the pressure at the exit cross-section is greater than the back-pressure aP_c , the actual velocity is less than this, and, with the same assumptions, will be

$$u_b = \sqrt{2g \cdot \frac{n}{n-1} \cdot P_1 V_1 \cdot \left(1 - b^{\frac{n-1}{n}} \cdot \frac{2}{n+1} \right)}.$$

Assuming first of all that the jet can gain no further velocity after it has left the nozzle, the loss of energy due to the under-expansion is

$$\begin{aligned} \Delta E_1 &= 1 - \frac{u_b^2}{u_a^2} \\ &= \frac{2}{n+1} \frac{\left(b^{\frac{n-1}{n}} - a^{\frac{n-1}{n}} \right)}{1 - \frac{2}{n+1} \cdot a^{\frac{n-1}{n}}}. \quad \dots \quad (1) \end{aligned}$$

The amount of this energy loss is shown plotted in fig. 1, covering a range of values of a and b . It will be seen that the losses can be serious, particularly when a is small and b approaches unity.

The losses shown in fig. 1 probably represent the worst possible condition, since the kinetic energy available from the expanded fluid is supposed to be only that with which it is discharged. At least some part of the jet, however, will have an opportunity to gain further kinetic energy after it has left the nozzle.

To examine the effect of this, consider the case of a simple converging nozzle discharging into a region where the pressure is lower than the critical pressure. The jet is supposed to leave the nozzle with the critical velocity u_c and its cross-sectional area is A . As before, let P_c equal the throat (critical) pressure and P_2 the external pressure at the discharge end. Then, writing down the equation of momentum for that

section of the jet where the mean pressure equals the external pressure, we have

$$\frac{W}{g}(u-u_e)=(P_e-P_2)A,$$

where $\frac{W}{g}$ is the mass accelerated per second and u is the mean velocity over the section.

But

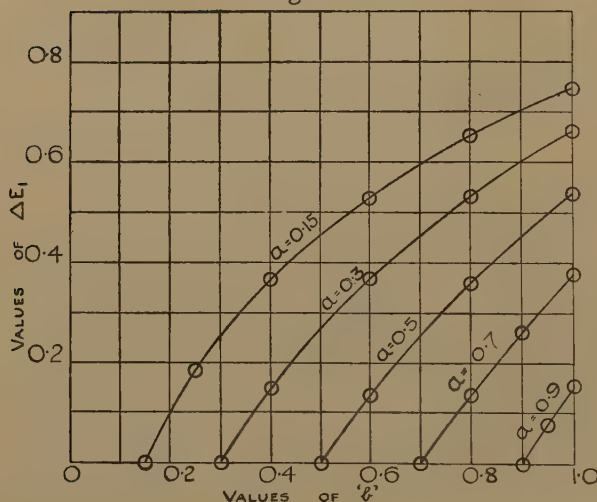
$$u_e=\sqrt{2g\frac{n}{n+1}P_1V_1},$$

$$P_e=\left(\frac{2}{n+1}\right)^{\frac{n}{n-1}}P_1,$$

$$W=\frac{Au_e}{\left(\frac{P_1}{P_e}\right)^{\frac{1}{n}}V_1},$$

$$P_2=aP_e;$$

Fig. 1.



hence, performing the various substitutions, we obtain, after a little reduction,

$$u=\sqrt{P_1V_1}\left[\sqrt{2g\frac{n}{n+1}}+\frac{g(1-a)\frac{2}{n+1}}{\sqrt{2g\frac{n}{n+1}}}\right];$$

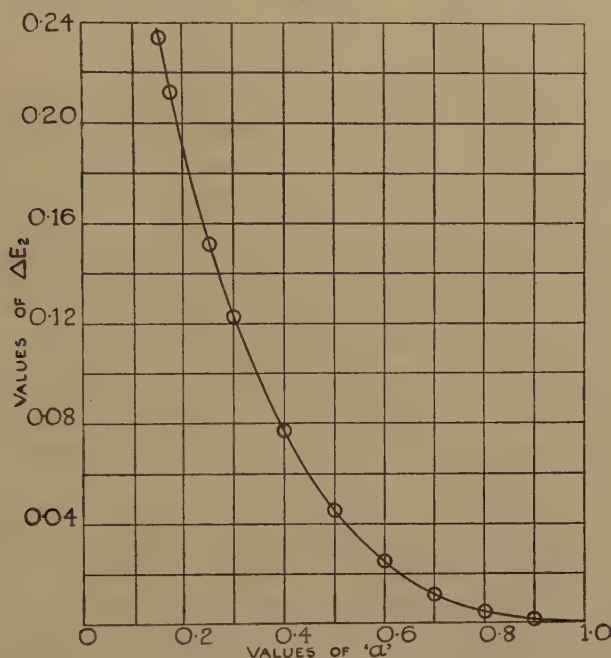
so in this case the energy loss

$$\begin{aligned}\Delta E_2 &= 1 - \frac{u^2}{u_a^2} \\ &= 1 - \frac{\left\{\sqrt{2g\frac{n}{n+1}} + \frac{g(1-a)\frac{2}{n+1}}{\sqrt{2g\frac{n}{n+1}}}\right\}^2}{\left\{\sqrt{2g\frac{n}{n-1}}\left[1-a^{\frac{n-1}{n}}\cdot\frac{2}{n+1}\right]\right\}^2}\end{aligned}$$

$$= 1 - \frac{n-1}{n^2(n+1)} \cdot \frac{(n+1-a)^2}{1 - \frac{2}{n+1} \cdot a^{\frac{n-1}{n}}} \quad \dots \quad (2)$$

Values of ΔE_2 are shown plotted in fig. 2. The loss, for any particular value of the constant a , may be compared with that given in fig. 1 for the same value of a , the constant b in this case being always unity, since the nozzle has no diverging portion. It will be observed that the loss is very considerably reduced when the jet is allowed to gain velocity after leaving the nozzle, and this should also apply when the

Fig. 2.



nozzle has a divergent extension after the throat, but when the exit area of such an extension is insufficient for the full range of expansion to take place within the nozzle.

In practice, the amount of the increase in velocity which may occur after the under-expanded jet has left the nozzle depends upon the conditions obtaining in any given case. Experimental results* suggest that the kinetic energy of the jet tends to increase with distance from the end of the nozzle; hence, if the jet is made to do work as soon as it leaves the nozzle, the losses due to under-expansion may be relatively high, though it would seem unlikely that they would ever reach the limiting values of fig. 1.

* E. Lewicki, *Z. Ver. deutsch. Ing.* p. 441 (1903).

It has been observed that an interesting example of under-expansion may occur, under normal working conditions, with the "exhaust steam injector," a form of boiler-feed pump employing low-pressure exhaust steam as the source of energy. In one form of the apparatus the steam supply is admitted in two stages, the major part, or "primary steam," entering through a central nozzle, at the discharge end of which it is brought into intimate contact with an annular jet of water to which it transfers momentum and by which it is condensed. The remainder of the steam—the "secondary" supply—then gives a further impetus to the high-velocity water jet formed by the union of the primary supply and the condensing water. After condensation of the whole of the steam has been completed, part of the kinetic energy of the hot-water jet is changed into pressure energy in a diffuser, in order that the resistance of the boiler pressure may be overcome.

In a case investigated by the present writer, the steam pressure at admission was 16 lb. per square inch (absolute), the temperature of the water in the supply tank 60° F., and the overall mass ratio, water to steam, was 10 : 1. The proportions by weight of primary and secondary steam were 67 per cent. and 33 per cent. respectively, so that at the discharge end of the primary steam nozzle—with which we are concerned in the subsequent discussion—the effective water-steam ratio was $\frac{10}{2/3} = 15 : 1$. The mean pressure in the region into which the primary steam-nozzle discharged was found to be 2.25 lb. per square inch (absolute), and the diameters at the throat and exit cross-sections of the nozzle were 24 mm. and 26 mm. respectively, giving an area ratio of 0.852 to 1.

On calculating the throat and exit areas of a nozzle to expand initially dry saturated steam from 16 lb. per square inch (absolute) to 2.25 lb. per square inch (absolute), and assuming a 5 per cent. energy loss * due to friction, it is found that the ratio of throat to exit area should be approximately 0.513 to 1. The use of the steam-nozzle of the dimensions quoted above therefore results in a considerable degree of under-expansion.

A fully-expanding steam-nozzle was designed and made with the calculated throat-exit area ratio, the absolute throat area remaining 24 mm. as before. This nozzle was substituted for the first, all other conditions remaining unchanged, and its effect on the performance of the apparatus was investigated by throttling the hot-water supply delivered by the diffuser, and by measuring the greatest pressure against which the injector would deliver without wastage of water. By comparing the results obtained with each of the two nozzles, and making use of the knowledge obtained in another series of experiments †, of the

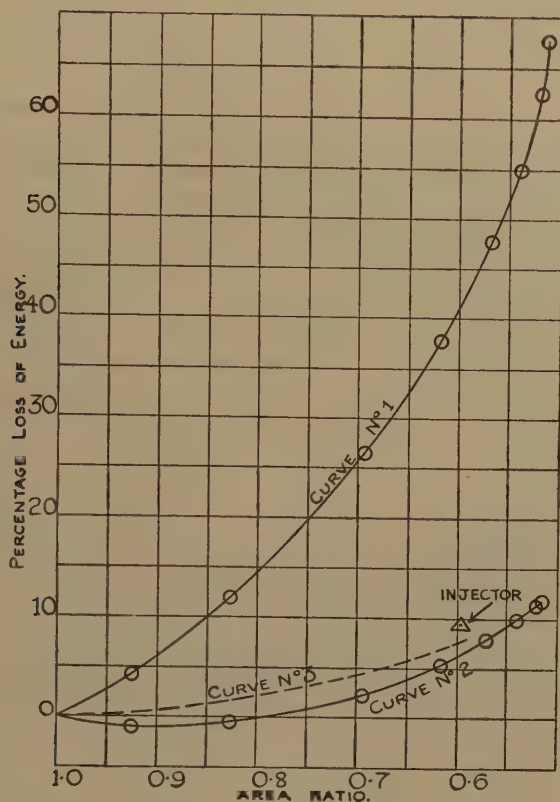
* Since with injectors the entrance passage to the throat is commonly relatively long, friction is supposed—contrary to the usual assumption—to take effect before as well as after the throat, resulting in an increase in both throat and exit areas relative to the condition with no friction.

† L. Kastner, *Journ. Inst. Loco. Engineers*, no. 147, vol xxix. p. 273.

losses occurring during the mixing process, it was estimated that the use of the under-expanding nozzle resulted in an energy loss of some $9\frac{1}{2}$ per cent. in the primary steam jet, relative to the performance when the fully-expanding nozzle was employed.

In the case discussed, the outer surface of the steam jet will come into contact with the annular water jet as soon as it issues from the nozzle, and will be rapidly condensed, whilst a central core of steam will probably travel some distance before suffering condensation. The proportion of the steam which can attain the full increase in velocity which the excess

Fig. 3.



pressure could produce can hardly be estimated, but the fact that the energy loss due to under-expansion is relatively so small in such an extreme case appears to support the view that, on the average, the final velocity attained is very considerably higher than that with which the jet leaves the nozzle, and that the increase in velocity takes effect within a small distance.

Fig. 3 shows the loss of energy due to under-expansion plotted against the ratio actual exit area divided by exit area for complete expansion. Curve no. 1 has been calculated, with the help of a large-scale Mollier

total heat-entropy chart, for steam expanding from 16 lb. per square inch (absolute) to 2.25 lb. per square inch (absolute), with a nozzle friction allowance of 5 per cent., and roughly corresponds to the limiting case of fig. 1, when $a = \frac{2.25}{0.577 \times 16} = 0.244$. Curve no. 2 has been obtained

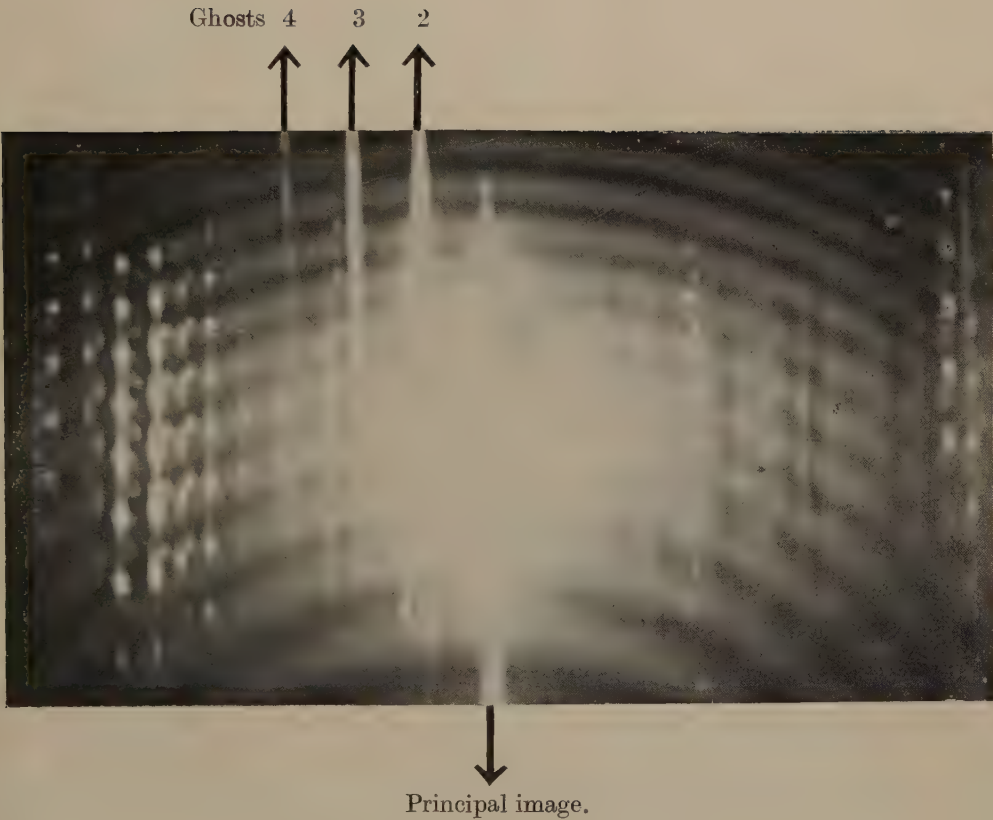
for the same pressure ratio by extending the use of equation (2) to cases where the nozzle is convergent-divergent in form but deficient in exit area, and where the steam, after expanding in a nozzle in which it suffers a 5 per cent. friction loss, gains further velocity after leaving the nozzle.

Although such an extension of equation (2) may be open to question, it has been thought justifiable to employ it in the present instance, substituting for u_c the velocity at the exit cross-section of the under-expanding convergent-divergent nozzle, and for A and p_c the area and pressure at the discharge end of such a nozzle. The results of this substitution, shown plotted in curve no. 2, may be compared with curve no. 3, which embodies certain figures relating to under-expansion given by Steinmetz *. The curves nos. 2 and 3 show closely similar trends. It will be observed that curve no. 2 suggests that a small degree of under-expansion—corresponding, say, to an area ratio of 0.9 or so—may sometimes not be undesirable, but may even be of some slight advantage. This is logical, and is to be expected, since whenever there is friction in the nozzle a stage may be reached after which it is more profitable to complete the expansion outside the nozzle, the losses due to friction in the passage then exceeding those due to free expansion.

The point marked with an arrow refers to under-expansion in the primary steam nozzle of the exhaust steam injector, and, although the loss is here somewhat greater than that shown by curves 2 and 3, it is still not excessive.

It is concluded that the losses due to under-expansion in a nozzle are seldom likely to be serious, even when the ratio between the admission pressure and the discharge pressure is large, and the cross-sectional area of the nozzle at exit is considerably less than its proper value. If, after leaving the nozzle with a pressure higher than the external pressure, the fluid were unable to gain any further kinetic energy, calculation shows that the losses would often be excessive; but, happily, this limiting case is not found to occur in practice, and the bulk of the fluid is rapidly accelerated by the excess pressure, and may reach a velocity not very far short of the value it would attain in a smooth nozzle designed to allow for the full range of expansion.

* E. Steinmetz, Proc. Amer. Soc. Mech. Eng. vol. xxx. no. 5, p. 628 (May 1908).



XXXV. *Methods for the Investigation of the Properties of Liquids at Ultra High Frequencies.*

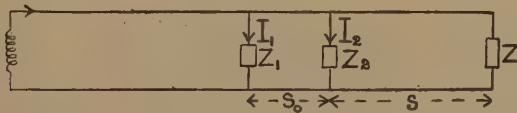
By GORDON WILLIAMS, M.Sc., Ph.D., F.Inst.P., Wheatstone Laboratory, King's College, London *.

[Received February 4, 1944.]

IN a recent paper ⁽¹⁾ the author developed a method for the measurement of impedance at ultra high frequencies, using Lecher wires or coaxial tubes, in which errors due to variation in the power output of the oscillator supplying the energy are eliminated. The Lecher wires are terminated by the unknown impedance whose value is calculated from the ratios of the currents flowing in two other bridges placed in various positions along the wires. The bridges can be moved independently, or together in the form of a tandem bridge. The latter arrangement is more convenient, and, provided the separation of the bridges is adjusted to one of a series of fixed values, a simple expression is obtained for the ratio of the currents, from which the components of the impedance can be computed.

The arrangement of the bridges is shown in fig. 1. Z_1 and Z_2 are suitable current meters and Z is the unknown impedance.

Fig. 1.



When the separation of the bridges, s_0 , is such that

$$\tan \frac{2\pi}{\lambda} s_0 = -\frac{Z_2^2}{Z_0 X_2}, \quad \dots \dots \dots (1)$$

where λ is the wave-length of the oscillations, Z_0 is the characteristic impedance of the wires, and X_2 is the reactive component of Z_2 , then

$$\rho^2 = \left| \frac{I_1}{I_2} \right|^2 = K_1 + \frac{K_2}{\sinh^2 a + \sin^2 \left(b + \frac{2\pi}{\lambda} s \right)}. \quad \dots \dots \dots (2)$$

The quantities, a and b , are related to Z by the equation

$$\frac{Z_0 - Z}{Z_0 + Z} = e^{-2(a+jb)}. \quad \dots \dots \dots (3)$$

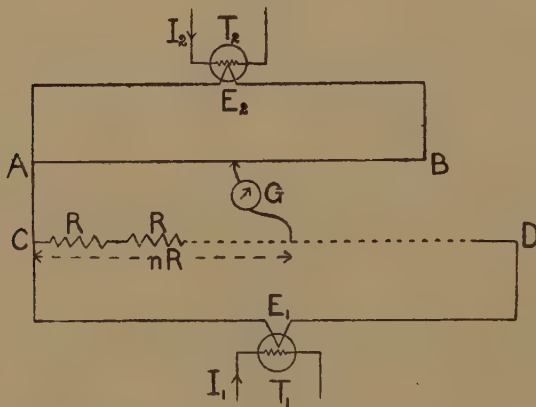
* Communicated by Dr. H. T. Flint.

For a given wave-length, K_1 is a function of the line parameters only, whereas K_2 is also a function of a . The value of K_2 , however, is not required.

In the reasoning leading to the results quoted above, the characteristic impedance of the wires was assumed to be the same at all points. It is, however, not always possible when constructing apparatus to make the characteristic impedance between the bridges the same as that between Z_2 and Z , due to the fact that the bridges have to be rigidly connected together. The theory can, however, be modified to include this case.

Let the characteristic impedance between the tandem bridges be real, and be designated by Z'_0 . If this is substituted in the appropriate equations in the original paper, the critical values of s_0 are given by an equation identical with equation (1), but with Z'_0 taking the place of Z_0 . Since, however, the correct separation will be found experimentally, the

Fig. 2.



value of Z'_0 is not required. The final equations remain unchanged. If the losses along the line between the bridges are appreciable, that is, if Z'_0 is complex, the correct values of s_0 cannot easily be ascertained. This is of course avoided by suitable construction of the apparatus.

Use of the Method.

Measurement of ρ^2 .

The currents I_1 and I_2 are conveniently measured using vacuum thermojunctions connected to d -c microammeters. The e.m.f. produced in each thermocouple is proportional to the square of the heater current, so that the ratios of the microammeter readings give the values $k\rho^2$, where k is a constant independent of the magnitude of the current involved, but is a function of frequency ⁽²⁾.

A more direct method of determining $k\rho^2$, which is the quantity used in the calculations, is to use the circuit shown in fig. 2,

Thermojunction T_2 is used to supply e.m.f. for the potentiometer AB. The output of the other thermojunction, T_1 , is connected to a series of equal resistances, CD, so that any known fraction of the potential difference across CD can be tapped off.

When balance is obtained on the potentiometer,

$$\frac{E_1}{E_2} \propto \frac{l}{n},$$

that is

$$\rho^2 \propto \frac{l}{n}.$$

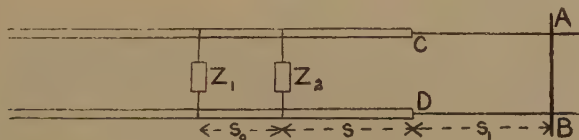
The absolute value of ρ^2 is not required ⁽¹⁾. The range of the instrument is altered by varying n .

Construction of Lecher Wires.

The method is most accurate when the impedance to be measured has approximately the same modulus as the characteristic impedance of the Lecher wires, so that in practice the wires are constructed to satisfy as nearly as possible this condition ⁽³⁾. In addition, the wires must be rigid and accurately fixed with a uniform separation. This frequently involves the use of lines whose characteristic impedance cannot be easily calculated, for example, lines constructed from lengths of angle brass. In this case, the characteristic impedance has to be determined experimentally by connecting a known impedance to the end of the line.

A suitable impedance consists of a subsidiary line of length s_1 constructed of two parallel cylinders short-circuited with a sliding plate, AB (fig. 3). The impedance of this line

Fig. 3.



is $jZ'_0 \tan \frac{2\pi}{\lambda} s_1$, where Z'_0 , its characteristic impedance, is easily calculable.

If s_1 is less than $\lambda/4$, this impedance is purely inductive, so that the first turning point in a graph of ρ^2 against s will be a minimum. Only the position of this turning point is required. If s' is its distance from CD, the end of the measuring lines, the impedance of the subsidiary line will be

$$jZ'_0 \tan b, \text{ where } b = \frac{\pi}{2} - \frac{2\pi}{\lambda} s'.$$

Hence

$$\begin{aligned} Z_0 &= Z'_0 \frac{\tan \frac{2\pi}{\lambda} s_1}{\tan b} \\ &= Z'_0 \tan \frac{2\pi}{\lambda} s_1 \tan \frac{2\pi}{\lambda} s'. \end{aligned} \quad \dots \dots (4)$$

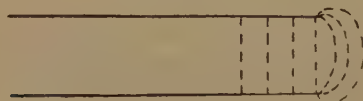
Determination of s_0 and K_1 .

For a given apparatus the values of s_0 and K_1 vary only with wave-length. Thus, to avoid the necessity of determining these quantities each time the wave-length is changed, it is convenient to determine their values for a number of wave-lengths⁽¹⁾ and to plot graphs of s_0 against λ , and K_1 against λ . The graphs obtained are standard graphs for the apparatus, from which the values of s_0 and K_1 for any wave-length can be obtained. The graphs are especially useful when the variation of the value of an impedance with wave-length is being investigated, as there will be no need to remove the impedance from the end of the Lecher wires during the investigation.

Elimination of End Effect.

The stray displacement currents which exist beyond the end of the Lecher wires (fig. 4) introduce an end effect which is equivalent to the presence of a small capacity situated at the end. They constitute a source of error, the magnitude of which depends upon the geometry and value of the impedance being measured. By suitable design of the

Fig. 4.



terminating impedance the error can be made small, but, as in most cases the impedance is of a fixed size and shape, the end effect may have to be eliminated.

One method is to attach to the end of the Lecher wires a short-circuited subsidiary line of known characteristic impedance. This will be in parallel with the unknown impedance whose value is calculated, using the measured value of the total end impedance. This arrangement reduces the end effect since no displacement currents exist beyond the short-circuit.

When the length of the subsidiary line is equal to $\lambda/4$ its impedance becomes infinite. The conditions existing at the end of the Lecher wires are then the same as would exist without the subsidiary line and with no end effect.

Alternatively, the reactance of the subsidiary line can be adjusted to compensate completely for the reactance of the unknown impedance, when the total impedance becomes entirely resistive. Under these conditions the first minimum occurs at a value of s equal to $\lambda/4$.

A further modification is to short-circuit the Lecher wires themselves and to place the unknown impedance, Z , at some other point, s being measured from Z (fig. 5).

Investigation of the Properties of Liquids.

(1) Using Lecher Wires.

A small condenser is attached to the Lecher wires and its capacity, C_0 , in air is determined. The fact that this will not be the same as the calculated value of the capacity due to edge effects will not affect the results, since the plates will be well immersed in the liquid under investigation.

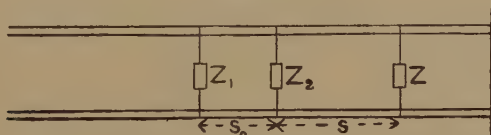
The cell is then filled with the liquid and the series reactive and resistive components of the liquid condenser so formed are determined. The equivalent parallel capacity, C_L , and resistance, R_L , are calculated from these values.

The dielectric constant of the liquid will then be $\epsilon = \frac{C_L}{C_0}$, and its specific conductivity will be

$$\sigma = \frac{1}{4\pi C_0 R_L 9 \cdot 10^{11}} \text{ ohm}^{-1}\text{-cm.}^{-1}$$

C_0 will be in farads and R_L in ohms.

Fig. 5.



The above formula for σ is quite general and applies to condensers of any shape and orientation⁽⁴⁾.

If the errors associated with the condenser cell itself are large, sets of results should be taken for two separations of the condenser plates.

(2) Using Coaxial Lines.

Method (i).—In the first method to be described the centre conductor is terminated a short distance from the end of the outer tube, which is closed by a conducting plate P (fig. 6). The arrangement is then equivalent to a small condenser attached to the line at some point A.

Distances are measured from the point A, and the properties of the liquid are determined as in the previous case.

Both the methods so far described have the advantage that only a small quantity of liquid is required.

Method (ii).—In this method the impedance of a known length of "shorted" coaxial line is measured, with and without liquid between the conductors. The most convenient way of doing this is to short-circuit the coaxial line itself, and to use the portion of it between the end and some point A (fig. 7).

There are two cases to be considered. In the first the losses in the liquid are so small that they are negligible, whilst in the second they are comparatively large and have to be taken into account.

(a) *Liquids in which losses are negligible.*

In this case the impedance of the liquid-filled line becomes a pure reactance having the value $Z'_0 \tan \beta_1 s_1$, where Z'_0 , the characteristic impedance of the line, is real, and β_1 , the wave-length constant, is equal to $\frac{2\pi}{\lambda_1}$, λ_1 being the wave-length of the oscillations in the liquid. This reactance becomes infinite when

$$\frac{2\pi}{\lambda_1} s_1 = \frac{\pi}{2}, \frac{3\pi}{2}, \text{ etc., that is, when } s_1 = \frac{\lambda_1}{4}, \frac{3\lambda_1}{4}, \text{ etc.,}$$

and zero when $\frac{2\pi}{\lambda_1} s_1 = 0, \pi, \text{ etc., that is, when } s_1 = 0, \frac{\lambda_1}{2}, \text{ etc.}$

Fig. 6.

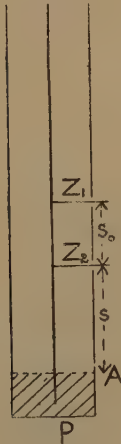
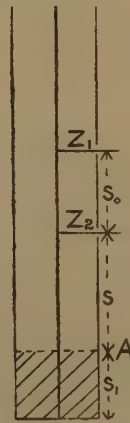


Fig. 7.



Thus, by varying the length of the liquid-filled line and measuring its impedance, the value of λ_1 can be accurately found. The impedance beyond A is infinite when the first maximum of ρ^2 occurs for a value of s equal to $\lambda/4$, λ being the wave-length of the oscillations in air. The impedance is zero when the first minimum of ρ^2 occurs when s is equal to $\lambda/4$. The dielectric constant of the liquid can then be calculated from the formula

$$\epsilon = \left(\frac{\lambda}{\lambda_1} \right)^2.$$

(b) *Liquids in which Losses cannot be Neglected.*

If the method is to be applied in this case, suitable expressions for the losses and the dielectric constant of the liquid have to be obtained in terms of the quantities actually measured, namely, the series resistive

and reactive components of the liquid-filled line and the wave-length constant β_1 .

Let the propagation constant of the liquid-filled line be P_1 . This will be complex, and can be written in the form $\alpha_1 + j\beta_1$, where α_1 is the attenuation constant of the line. The values of α_1 and β_1 are given by the usual expressions

$$\alpha_1 = \sqrt{\frac{1}{2} \{ \sqrt{(\text{GR} + \text{LC}\omega^2)^2 + \omega^2(\text{LG} - \text{CR})^2} + (\text{GR} - \text{LC}\omega^2) \}} \quad . \quad . \quad (5)$$

and
$$\beta_1 = \sqrt{\frac{1}{2} \{ \sqrt{(\text{GR} + \text{LC}\omega^2)^2 + \omega^2(\text{LG} - \text{CR})^2} - (\text{GR} - \text{LC}\omega^2) \}}, \quad . \quad . \quad (6)$$

where R is the resistance, G is the dielectric conductance, L is the inductance, and C is the capacity of the line, all per unit length of lead and return.

In the present case R is negligible, whilst L has the same value as for the air-filled line and can be calculated. G and C are the quantities required. Their values can be calculated if α_1 and β_1 are known. The conductivity of the liquid is then obtained from G, and the dielectric constant, ϵ , from C and C_0 , C_0 being the capacity per unit loop length of the line in air.

On putting $R=0$ in expressions (5) and (6) we find that

$$\alpha_1 = \sqrt{\frac{1}{2} \text{L}\omega \{ \sqrt{\text{C}^2\omega^2 + \text{G}^2} - \text{C}\omega \}} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and
$$\beta_1 = \sqrt{\frac{1}{2} \text{L}\omega \{ \sqrt{\text{C}^2\omega^2 + \text{G}^2} + \text{C}\omega \}}. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Solving these equations for C and G we obtain the equations

$$\alpha_1^2 - \beta_1^2 = -\text{LC}\omega^2 = -\text{L}\epsilon\text{C}_0\omega^2 \quad . \quad . \quad . \quad . \quad . \quad (9)$$

and
$$2\alpha_1\beta_1 = \text{LG}\omega. \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Determination of α_1 and β_1 .

Let the value of the terminating impedance as measured be $R_1 + jX_1$. This is equal to $Z'_0 \tanh (\alpha_1 + j\beta_1)s_1$, where s_1 is the length of the "shorted" liquid-filled line, and Z'_0 is its characteristic impedance, which can be written $R'_0 + jX'_0$.

Thus

$$\begin{aligned} R_1 + jX_1 &= (R'_0 + jX'_0) \tanh (\alpha_1 + j\beta_1)s_1 \\ &= (R'_0 + jX'_0) \frac{\sinh 2\alpha_1s_1 + j \sin 2\beta_1s_1}{\cosh 2\alpha_1s_1 + \cos 2\beta_1s_1} \quad . \quad . \quad . \quad . \quad . \quad (11) \\ &= \frac{R'_0 \sinh 2\alpha_1s_1 - X'_0 \sin 2\beta_1s_1}{\cosh 2\alpha_1s_1 + \cos 2\beta_1s_1} + j \frac{R'_0 \sin 2\beta_1s_1 + X'_0 \sinh 2\alpha_1s_1}{\cosh 2\alpha_1s_1 + \cos 2\beta_1s_1}. \quad . \quad . \quad . \quad (12) \end{aligned}$$

Since α_1 , β_1 , R'_0 and X'_0 are all known functions of R, G, L and C, it should be possible to solve the above complex equation for G and C. However, the quantities are so arranged that, in practice, no satisfactory solution can be obtained, and another method has to be adopted. The following method is suggested.

Determination of β_1 .

β_1 can be calculated directly from the wave-length, λ_1 , of the oscillations in the liquid-filled line.

From equation (11) we have, equating moduli,

$$R_1^2 + X_1^2 = [(R'_0)^2 + (X'_0)^2] \left[\frac{\sinh^2 2\alpha_1 s_1 + \sin^2 2\beta_1 s_1}{(\cosh 2\alpha_1 s_1 + \cos 2\beta_1 s_1)^2} \right]. \quad (13)$$

In this equation R'_0 and X'_0 are independent of s_1 , whilst both $\sinh 2\alpha_1 s_1$ and $\cosh 2\alpha_1 s_1$ increase together with s_1 , so that, to a close approximation, the modulus of the impedance of the liquid-filled line will attain minimum values when

$$\cos 2\beta_1 s_1 = +1 \quad \text{and} \quad \sin 2\beta_1 s_1 = 0,$$

that is, when

$$2\beta_1 s_1 \left(= \frac{4\pi}{\lambda_1} s_1 \right) = 0, 2\pi, \text{ etc.},$$

or for values of s_1 equal to 0, $\frac{\lambda_1}{2}$, etc.

The accuracy of the value obtained for β_1 will depend upon the value of α_1 . However, since α_1 will eventually be calculated, the error involved can be determined by substitution in equation (13), and hence, a more accurate value of β_1 can be obtained. This process is then continued until the error becomes negligible.

The experimental procedure thus consists in measuring the impedances of known lengths of liquid-filled line, and finding the value of s_1 corresponding to the minimum value of $R_1^2 + X_1^2$.

Determination of α_1 .

Equating real and imaginary quantities in equation (12) we have

$$R_1 = \frac{R'_0 \sinh 2\alpha_1 s_1 - X'_0 \sin 2\beta_1 s_1}{\cosh 2\alpha_1 s_1 + \cos 2\beta_1 s_1} \quad \dots \quad (14)$$

and

$$X_1 = \frac{R'_0 \sin 2\beta_1 s_1 + X'_0 \sinh 2\alpha_1 s_1}{\cosh 2\alpha_1 s_1 + \cos 2\beta_1 s_1} \quad \dots \quad (15)$$

Therefore

$$\frac{R_1}{X_1} = \frac{R'_0 \sinh 2\alpha_1 s_1 - X'_0 \sin 2\beta_1 s_1}{R'_0 \sin 2\beta_1 s_1 + X'_0 \sinh 2\alpha_1 s_1}.$$

This equation can be rearranged to give

$$\frac{R_1 \sinh 2\alpha_1 s_1 + X_1 \sin 2\beta_1 s_1}{X_1 \sinh 2\alpha_1 s_1 - R_1 \sin 2\beta_1 s_1} = \frac{R'_0}{X'_0} = \text{constant}. \quad \dots \quad (16)$$

Hence α_1 can be determined by taking two lengths of line and measuring R_1 and X_1 in each case. It is much easier, however, to determine the value of the above constant in terms of α_1 and β_1 .

The characteristic impedance of the liquid-filled line, $R'_0 + jX'_0$, is equal to $\sqrt{\frac{R + jL\omega}{G + jC\omega}}$, so that when $R=0$ we can write

$$R'_0 + jX'_0 = \sqrt{\frac{jL\omega}{G + jC\omega}} = \sqrt{\frac{LC\omega^2 + jGL\omega}{G^2 + C^2\omega^2}}.$$

Equating the moduli of the above equation we have

$$(R'_0)^2 + (X'_0)^2 = \frac{\sqrt{L^2 C^2 \omega^4 + G^2 L^2 \omega^2}}{G^2 + C^2 \omega^2} \quad (17)$$

and equating real terms

$$(R'_0)^2 - (X'_0)^2 = \frac{LC\omega^2}{G^2 + C^2 \omega^2} \quad (18)$$

The values of R'_0 and X'_0 obtained from equations (17) and (18) are

$$R'_0 = \sqrt{\frac{1}{2} L \omega \left[\frac{\sqrt{C^2 \omega^2 + G^2} + C \omega}{G^2 + C^2 \omega^2} \right]},$$

and

$$X'_0 = \sqrt{\frac{1}{2} L \omega \left[\frac{\sqrt{C^2 \omega^2 + G^2} - C \omega}{G^2 + C^2 \omega^2} \right]},$$

so that

$$\frac{R'_0}{X'_0} = \frac{\sqrt{[\sqrt{C^2 \omega^2 + G^2} + C \omega]}}{\sqrt{[\sqrt{C^2 \omega^2 + G^2} - C \omega]}} \quad (19)$$

It is seen from equations (7) and (8) that the quantity on the right-hand side of the above equation is equal to $\frac{\beta_1}{\alpha_1}$, so that

$$\frac{R'_0}{X'_0} = \frac{\beta_1}{\alpha_1}, \quad (20)$$

a relationship which holds only when $R=0$.

Equation (16) can therefore be written in the form

$$\frac{R_1 \sinh 2\alpha_1 s_1 + X_1 \sin 2\beta_1 s_1}{X_1 \sinh 2\alpha_1 s_1 - R_1 \sin 2\beta_1 s_1} = \frac{\beta_1}{\alpha_1} \quad (21)$$

This can be rearranged to give the equation

$$\sinh 2\alpha_1 s_1 = \left[\frac{\alpha_1 + \beta_1 \frac{R_1}{X_1}}{\beta_1 - \alpha_1 \frac{R_1}{X_1}} \right] \sin 2\beta_1 s_1 \quad (22)$$

This equation can readily be solved by plotting both the left-hand side and the right-hand side against α_1 and obtaining the intersection of the two curves. ϵ and G can then be calculated, using equations (9) and (10).

The procedure is therefore summarized as follows:—First obtain the value of β_1 as described. Then use the experimental values of R_1 and X_1 , obtained for a known length of liquid-filled line, to calculate α_1 . If necessary, this value of α_1 can be used to deduce a more accurate value of β_1 . Suitable values of R_1 and X_1 will have been obtained during the first determination of β_1 .

A more convenient way of determining ϵ and G is to determine the values of R_1 and X_1 for lengths of liquid-filled line equal to $\frac{\lambda_1}{4}$, $\frac{\lambda_1}{2}$, etc.

For these lengths $\sin 2\beta_1 s_1 = 0$, so that we have from equation (22)

$$\frac{\beta_1}{\alpha_1} = \frac{R_1}{X_1} = \text{constant},$$

that is,
$$\alpha_1 = \frac{X_1}{R_1} \beta_1 = \frac{X_1}{R_1} \cdot \frac{2\pi}{\lambda_1} \dots \dots \dots (23)$$

Substituting this value of α_1 in equation (9) we obtain the following expression for the dielectric constant,

$$\begin{aligned} \epsilon &= \left[1 - \frac{X_1^2}{R_1^2} \right] \frac{\beta_1^2}{LC_0 \omega^2} \\ &= \left[1 - \frac{X_1^2}{R_1^2} \right] \frac{1}{LC_0 (n\lambda_1)^2}, \dots \dots \dots (24) \end{aligned}$$

where n is the frequency of the oscillations.

But $n = c/\lambda$, where c is the velocity of light and λ is the wave-length of the oscillations in free space, that is in the air-filled line, so that we have

$$\begin{aligned} \epsilon &= \left[1 - \frac{X_1^2}{R_1^2} \right] \frac{1}{LC_0 c^2} \cdot \left(\frac{\lambda}{\lambda_1} \right)^2 \\ &= \left[1 - \frac{X_1^2}{R_1^2} \right] \left(\frac{\lambda}{\lambda_1} \right)^2, \text{ since } \frac{1}{LC_0} = c^2. \dots \dots \dots (25) \end{aligned}$$

Substituting in equation (10) we obtain

$$G = 2 \frac{X_1}{R_1} \cdot \frac{\beta_1^2}{L\omega} \dots \dots \dots (26)$$

$$\begin{aligned} &= \frac{X_1}{R_1} \cdot \frac{4\pi}{Lc} \cdot \frac{\lambda}{\lambda_1^2} \\ &= \frac{X_1}{R_1} 4\pi C_0 c \frac{\lambda}{\lambda_1^2} \dots \dots \dots (27) \end{aligned}$$

Alternatively, if the values of R_1 and X_1 , measured for lengths of liquid-filled line equal to $\frac{\lambda_1}{4}$ and $\frac{\lambda_1}{2}$, are R'_1 and X'_1 , R''_1 and X''_1 respectively, we have, using equations (14) and (15),

$$\frac{R'_1}{R''_1} = \frac{X'_1}{X''_1} = \frac{\coth \alpha_1 \frac{\lambda_1}{4}}{\tanh \alpha_1 \frac{\lambda_1}{2}}, \dots \dots \dots (28)$$

from which α_1 may be calculated.

Method (iii).—Another method which appears to warrant consideration is that in which the whole of the coaxial tube is filled with the liquid under investigation. We shall discuss the method as it applies first to a loss-free liquid, and then to a liquid in which losses are not negligible.

(a) *Loss-free Liquid.*

The theory in this case is identical with that given in the original paper, so that the distance between consecutive maxima and minima

of ρ^2 will be $\frac{\lambda_1}{4}$, λ_1 again being the wave-length of the oscillations in the liquid. λ_1 can therefore be determined accurately. If λ is the corresponding wave-length in air, the dielectric constant of the liquid will be $\epsilon = \left(\frac{\lambda}{\lambda_1}\right)^2$.

(b) *Liquids in which Losses Cannot be Neglected.*

When the losses in the liquid are taken into account the final expression for ρ^2 , which, in the original theory is quite simple, becomes much too complicated to be of practical use. It can be simplified by short-circuiting the end of the coaxial line and filling it with liquid only to the junction Z_2 , when the expression for ρ^2 becomes

$$\rho^2 = K_1 + \frac{K_2 \cosh 2\alpha_1 s_1 + K_3 \sinh 2\alpha_1 s_1 + K_4 \sin \beta_1 s_1}{\sinh^2 \alpha_1 s_1 + \sin^2 \beta_1 s_1}.$$

The K 's are constants, which, though independent of s , are yet dependent upon the liquid in the line. It is seen that this expression is much too involved to be of practical use.

APPENDIX.

Further Methods for the Measurement of Impedance at Ultra High Frequencies.

Flint and Williams (1941)⁽⁵⁾ and Williams (1944)⁽¹⁾ have developed methods for the measurement of impedance using Lecher wires, in which the effect of variations in the power output of the oscillator supplying energy to the wires is eliminated. In both methods the unknown impedance is connected to the end of the wires, and the ratios of the currents flowing in two bridges, connected across the wires in various positions, are used to calculate its value. The methods differ in the arrangement of the measuring bridges.

The following methods involve the measurement of constant fractions of the currents actually flowing at various points in the lines themselves, and the theoretical expressions have been deduced on the assumption that these currents can be measured without taking any power from the lines. A close approximation to this may be obtained by using a powerful oscillator in conjunction with small pick-up loops connected to sensitive vacuum thermo-junctions.

Method I.

In this method the current flowing in the line is compared with that flowing in the unknown impedance, Z , connected to the end of the line.

Let the current at the point A (fig. 8) be I_1 and that flowing into Z be I_2 .

Then
$$\frac{I_1}{I_2} = \cosh Ps + \frac{Z}{Z_0} \sinh Ps, \quad \dots \dots \dots (1)$$

where Z_0 is the characteristic impedance of the line and P is its propagation constant. Losses are assumed to be negligible, so that Z_0 is real and P is equal to $j\beta s$, where $\beta=2\pi/\lambda$, λ being the wave-length of the oscillations.

It is convenient here to introduce the current reflection coefficient $\frac{Z_0-Z}{Z_0+Z}=e^{-2r}$, where r , which will, in general, be complex, can be represented by $a+jb$.

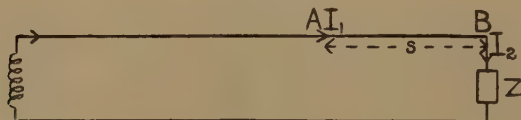
Hence $Z=Z_0 \tanh r$.

Substituting for Z in equation (1) gives the further equation

$$\begin{aligned}\frac{I_1}{I_2} &= \frac{1}{\cosh r} [\cosh r \cosh Ps + \sinh r \sinh Ps] \\ &= \frac{1}{\cosh r} \cdot \cosh (Ps+r) \\ &= \frac{1}{\cosh r} \cdot \cosh [a+j(b+\beta s)]. \quad \dots \dots \dots (2)\end{aligned}$$

In an experiment it is the modulus, ρ , of the ratio $\frac{I_1}{I_2}$, which is measured

Fig. 8.



The value of ρ^2 obtained from equation (2) is

$$\rho^2 = A[\sinh^2 a + \cos^2 (b+\beta s)] \quad \dots \dots \dots (3)$$

$$= A[\cosh 2a + \cos 2(b+\beta s)]/2, \quad \dots \dots \dots (4)$$

where $A = \frac{1}{|\cosh r|^2}$ is a constant which, though dependent upon Z , is independent of s .

Equations (3) and (4) are similar in form to the final equation obtained by Flint and Williams, so that they can be solved for a and b , using the method described in their paper.

The value of the resistive component of Z is

$$Z_0 \frac{\tanh a \sec^2 b}{1 + \tanh^2 a \tan^2 b} \quad \text{or} \quad Z_0 \frac{\sinh 2a}{\cosh 2a + \cos 2b}, \quad \dots \dots (5)$$

and that of the reactive component is

$$Z_0 \frac{\operatorname{sech}^2 a \tan b}{1 + \tanh^2 a \tan^2 b} \quad \text{or} \quad Z_0 \frac{\sin 2b}{\cosh 2a + \cos 2b}. \quad \dots \dots (6)$$

Method II.

In this method the currents flowing in the line at two points A and B (fig. 9) are measured simultaneously and their ratio is determined. The

positions of A and B are varied and the ratios obtained can be used to calculate the value of the unknown impedance connected to the end of the line. If the separation of A and B is fixed at a value equal to $\lambda/4$, a simple expression is obtained for the current ratio.

The ratio of the currents I_1 and I_2 is

$$\frac{I_1}{I_2} = \cosh Ps_0 + \frac{Z_r}{Z_0} \sinh Ps_0, \quad . \quad . \quad . \quad . \quad . \quad (7)$$

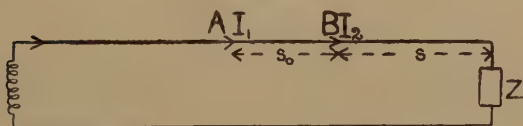
where Z_r is the total impedance beyond the point B. Using the same notation as in Method I, the expression for Z_r becomes

$$Z_0 \tanh [a + j(b + \beta s)].$$

On substituting this value in equation (7) we have

$$\begin{aligned} \frac{I_1}{I_2} &= \cosh Ps_0 \{1 + \tanh Ps_0 \tanh [a + j(b + \beta s)]\} \\ &= \cos \beta s_0 \{1 + j \tan \beta s_0 \tanh [a + j(b + \beta s)]\}, \quad . \quad . \quad . \quad (8) \end{aligned}$$

Fig. 9.



Again it is the modulus, ρ , of this expression which is measured in an experiment. The value of ρ^2 , obtained after separating real and imaginary quantities in equation (8), becomes

$$\rho^2 = 2 \cos^2 \beta s_0 - 1 + 2 \frac{\sin^2 \beta s_0 \cosh 2a - \sin \beta s_0 \cos \beta s_0 \sin 2(b + \beta s)}{\cosh 2a + \cos 2(b + \beta s)}. \quad (9)$$

This expression, which in its present form is too complicated to be of practical use, can be simplified by fixing s_0 at a value equal to $\lambda/4$, in which case $\beta s_0 = \pi/2$, so that $\cos \beta s_0 = 0$ and $\sin \beta s_0 = 1$. Equation (9) then takes the form

$$\rho^2 = -1 + 2 \frac{\cosh 2a}{\cosh 2a + \cos 2(b + \beta s)}. \quad . \quad . \quad . \quad . \quad (10)$$

It is impossible, at high frequencies, to measure the absolute value of ρ^2 , but a quantity ρ_1^2 equal to $k\rho^2$ may be determined, using pick-up loops, k being a constant for a fixed frequency. In practice equation (10) must therefore be written

$$\rho_1^2 = -k + 2k \frac{\cosh 2a}{\cosh 2a + \cos 2(b + \beta s)}. \quad . \quad . \quad . \quad . \quad (11)$$

This equation is similar to that obtained by Williams (1944)⁽¹⁾ and can therefore be solved using methods already described. It will be advisable to discuss in detail one of the methods suggested for the determination of a , since in this case the value of the constant k will not be required. Only the maximum and minimum values of ρ_1^2 , $\rho_{1(\max)}^2$ and $\rho_{1(\min)}^2$, respectively, are used,

From equation (11) it is seen that the maximum values of ρ_1^2 occur when $\cos 2(b+\beta s)=-1$ and the minimum values when $\cos 2(b+\beta s)=+1$, so that we can write,

$$\begin{aligned}\rho_{1(\max)}^2 &= k \left[-1 + \frac{\cosh^2 a + \sinh^2 a}{\sinh^2 a} \right] \\ &= k \coth^2 a \quad \dots \dots \dots (12)\end{aligned}$$

and

$$\begin{aligned}\rho_{1(\min)}^2 &= k \left[-1 + \frac{\cosh^2 a + \sinh^2 a}{\cosh^2 a} \right] \\ &= k \tanh^2 a. \quad \dots \dots \dots (13)\end{aligned}$$

These equations give

$$\frac{\rho_{1(\min)}^2}{\rho_{1(\max)}^2} = \tanh^4 a,$$

from which a can be calculated.

This method is advantageous for two reasons :

(a) it reduces the error in $\tanh a$ to approximately a quarter of the experimental error ;

(b) for large values of a , $\tanh a$ is nearly unity, but $\tanh^4 a$ will be considerably smaller and thus more readily ascertained experimentally.

Note added in proof.—In Method 1 of the Appendix, I_2 can represent the current at any point on the line without altering the form of the final expression for ρ^2 . In practice it would be convenient to measure I_2 at a point near the input end of the line.

References.

- (1) Williams, Proc. Phys. Soc. lvi. p. 63 (1944).
- (2) Williams and Rogers, 'Nature,' v. 149, p. 668 (June, 1942).
- (3) Rogers, Proc. Phys. Soc. lvi. p. 1 (1944).
- (4) Debye, 'Polar Molecules,' p. 99.
- (5) Flint, H. T., and Williams, G., Phil. Mag. xxxii. p. 489 (1941).

XXXVI. On the Dimensions of Physical Magnitudes (Fourth Paper).

By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.*

[Received June 1, 1944.]

In a recent issue of the Phil. Mag.†, Mr. D. P. Dalzell has made several criticisms of my paper on the dimensions of electric and magnetic magnitudes‡. I do not propose to comment on all of these individually,

* Communicated by the Author.

† Phil. Mag. xxxv. p. 188 (1944).

‡ *Ibid.* xxxiv. p. 588 (1943).

partly because some of them refer to details of the particular scheme I outlined, which I stated was merely an example and not the only legitimate one; though I am not convinced by Mr. Dalzell's arguments, I do not consider its defence worth the paper it would require. The main reason, however, is that Mr. Dalzell's article shows so clearly that he has missed the basic idea of my treatment of dimensions that it is far better to indicate the fundamental difference between our viewpoints, and leave the details to take care of themselves.

I regard our procedure in physics as the establishment of relations between the results of measurements. A measurement is a strictly defined operation resulting in a number. A straightforward description of the state of physics at any moment would therefore consist of (1) a list of the recognized measurements; (2) a statement, in the form of equations, of the experimentally determined relations between their results. I should have thought it obvious that one should seek to make (1) precise without assuming foreknowledge of any of the relations (2), but Mr. Dalzell calls such an endeavour a "vice." This is apparently because he is confusing it with a quite different ideal—namely, that of choosing the particular operations to be listed under (1) without regard to their power of leading to (2). Since (2) is an essential part of our aim, this ideal would indeed be vicious, and I have never aspired to it. An indefinitely large number of operations resulting in numbers could be imagined, and the few we elect to perform are those which experience tells us will give numbers which stand in simple relations to one another. For example, there is no reason why we should not allow each of various bodies to impinge on a standard system in a standard way, measure the duration of the sound produced, and call the result the "sonority" of the body. In itself that is no more fantastic than placing a tube of mercury in each of various bodies, noting the height to which the mercury rises, and calling the result the "temperature" of the body. We do not perform the first operation, and we frequently perform the second, and the reason is solely that temperatures are found to bear simple and important relations to the results of our other operations, and sonorities are not. Nevertheless, having included temperature in (1) we must then state the operation for determining its magnitude so that it could be carried out, with unambiguous results, by anyone without knowledge of the relations in (2).

Any operation of measurement can always include a final step in which a number already arrived at must be multiplied by a chosen numerical factor. The arbitrary choice of the factor is the arbitrary choice of a "unit." I called the factor N , and for simplicity supposed it to be unity in all the examples I gave. It is, of course, not always so. The operation of measuring length would be described by giving the various steps involved in comparing the body under test in a proper way with the standard metre, and adding that the result must be multiplied by 100 to get the value in terms of the generally accepted unit. Similarly, the operation of measuring a time interval would be

described by giving the various steps involved in determining in radians the corresponding change of hour angle of the mean Sun, and adding that the result must be multiplied by $\frac{86400}{2\pi}$ to get the value in terms

of the generally accepted unit. So far as principles are concerned, this final arbitrary step is utterly trivial. The dimensional symbol—L for length and T for time, for example—could stand throughout my treatment either for the whole operation of measurement or for that operation without this multiplication, on the understanding that a factor N, a “pure number,” can always be introduced if desired. The essential thing is the *whole operation* of measurement, not a trifling appendage of it which could be omitted with no disadvantage except a slight inconvenience.

Mr. Dalzell, however, writes throughout in terms of units as though they were vital. He may, of course, do this if he finds it helpful, but he misses the point entirely if he reads my papers with the presupposition that I share his experience. That he has done this seems evident from

his criticisms; for example, he states that I quote the equation $f = \frac{q_1 q_2}{\kappa r^2}$ “as though it were used for defining the electrostatic unit of charge.” I do nothing of the kind. I quote it as though it were used—as, in fact, it is—for specifying the kinds of operations used in measuring charge.

This misapprehension of Mr. Dalzell's is shown particularly clearly in his last sentence. He is referring to my example of a specification for measuring charge, which, for clearness, I quote: “*Choose any arbitrary standard charge, place the charge to be measured at a fixed position with respect to it, the whole system being in a vacuum, and measure the force between the charges; the magnitude of the unknown charge is then to be taken as proportional to this force.*” (“Proportional” instead of “equal,” of course, refers to the arbitrariness in the choice of unit.) On this Mr. Dalzell comments: “In Prof. Dingle's definition of charge one would naturally suppose that having assumed that an arbitrary standard charge can be chosen the simplest procedure would be to adopt it as a unit of comparison.” True, it would be simple, and absolutely useless. The problem before me was this: Here is a charge; state the operation for measuring it, *i. e.* for arriving at a number which represents its magnitude. My definition clearly enables one to do this. But what can Mr. Dalzell do? He knows only that some other charge is called a “unit.” How is he going to arrive at the number to be assigned to the charge which he is given? Obviously he will have to prescribe an operation for comparing two charges, and this operation must almost inevitably, in some respects at least, resemble mine. It will yield him a number, and in order that that number shall be consistent with the assignment of the number “1” to the standard charge, Mr. Dalzell will find himself bound to introduce a numerical factor N whose value he will not be able to control. In order, therefore, to make his “simplest

procedure" of any use, he has to add at least the equivalent of what I proposed. It is that addition that will do the work for him, and his original choice of a unit becomes simply an arbitrary restriction on his choice of N . That is the natural result of labouring trivial things like units instead of the essential operations of measurement.

One remark of Mr. Dalzell's, however, raises a point of some importance. He holds that charge, according to my definition quoted above, should be a fundamental and not a derived magnitude. The fact is that fundamental and derived magnitudes do not exhaust the whole scope of physical magnitudes. A simple operation, like the canonical operation for measuring length, is fundamental. A complex operation, consisting of a combination of fundamental magnitudes (*e. g.* the operation of measuring velocity, represented by L/T) is derived. But there is a third class which includes both; *e. g.* electric charge, the measurement of which includes the measurement of force and the independent operation of producing or selecting a standard charge and placing the unknown charge in a certain position relative to it. In my previous papers I have ignored the latter part of the operation when assigning the dimensional symbol, and represented the measurement of charge by the single symbol F . I think it is probably desirable that the symbolism should be modified to give separate recognition to this third class. I have not previously attempted this, being desirous of making as few innovations as possible, but I think this might have been a mistake. The difficulty would not be met, however, by ignoring the dependence of the measurement on force, and calling charge a fundamental magnitude, as Mr. Dalzell suggests. If charge is to be either fundamental or derived, I think its allocation to the derived category is distinctly the lesser evil.

Mr. Dalzell asks what I mean by calling the electrostatic and electromagnetic systems of measurement "equivalent." I hoped that, in its context, this would be clear, but I may add that I meant simply that the two systems of measurement covered the same set of operations and experimental relations, but these were distributed differently between the classes (1) and (2) referred to above. I gave a simple example of how this was possible in mechanics, on pp. 589-90 of the paper. Perhaps I may add that I did not mean that the systems were identical, or would necessarily always remain equivalent (I said they were equivalent "to the best of our knowledge"). A new effect of electric currents, for example, might be discovered which might be found not to result from "electrostatic" charges in motion. In that case the equivalence would break down. But at present I believe it is complete. The systems are alternatives; they are not necessary to supplement one another.

I fear I do not understand Mr. Dalzell's deduction that my scheme has no place for phenomena such as the Hall effect, etc. My point was simply that every operation in (1) should, in an absolutely rigorous treatment, be given a distinctive name, and the identity of the results obtained by certain different operations (*e. g.* operations with a measuring rod and by triangulation for, as we say, measuring "length") should

be recognized to be an experimentally established relation and come under (2). All the phenomena Mr. Dalzell mentions would naturally come under (2). Why should he imagine I must "disregard" them?

In general, I find nothing in Mr. Dalzell's paper which leads me to doubt the soundness of my view of dimensions or the validity of the scheme for electric and magnetic magnitudes by which I exemplified it. Nor can I see that he has met, as he set out to do, the objections to the generally-accepted scheme. The dimensions of κ and μ are generally stated to be "unknown." Mr. Dalzell has neither justified this—*i. e.* justified a mystical view of dimensions which regards them as objectively existing characteristics of nature which we are potentially able to "discover"—nor given, with reasons for the assignment, the dimensions which he thinks these quantities must possess. Until he has done one of these things, he has done nothing to justify the scheme which he supports.

Finally, I fully share Mr. Dalzell's desire for an unprejudiced approach to Eddington's theory, but I cannot agree with his criticism of my remarks thereon. I did not invent new dimensions in order to say that the fine-structure constant was not dimensionless in the form given. I took the dimensions which are given in any standard textbook, or in "Kaye and Laby," and they show, as Mr. Dalzell must know, that since e is the only one of the quantities in the constant which involves κ , the constant is not free from "our artificial standards." I do not know what Mr. Dalzell means by saying that "too much attention has been paid to Eddington's presentation of his work and too little attention paid to the work itself." So far as I know, the only possible approach to the work is through Eddington's presentation of it. If Sir Arthur Eddington's experience is similar to mine he will not echo Mr. Dalzell's complaint on his behalf. He will rather regret that too much comment has been made on his work with too little attention to his presentation of it.

XXXVII. *Problems of the Transmission of Light through Tissues and some other Media.*

By DAVID S. EVANS, M.A., Ph.D.*

[Received February 14, 1944.]

1. *Introduction.*

THE transmission of light through living tissues has recently aroused a certain interest, since the measurement of transmitted intensities has been made use of in several instruments designed to measure the oxygen

* Communicated by Dr. K. Mendelssohn.

content of blood (Squire (1940), Goldie (1942), Millikan (1942)). However, the problem is one which has many applications in physiological and biological investigations, and may also lead, as we shall see, to considerations of a more general character having a possible application to materials other than living tissues.

In this paper we shall deal only with the physical problem, which consists in an investigation of the transmission of light through a medium of two constituents, the second of which is distributed throughout the first in the form of a series of vessels. These vessels will all have some common general character, but will be subject to variations in size, shape and distribution.

The practical application in the case of blood oxygen meters consists in the fact that all these instruments depend on the differences in absorption spectrum between the blood-pigment, hæmoglobin, and the oxygen-bearing form of this substance, oxy-hæmoglobin. If, therefore, the absorption of a layer of blood-bearing tissues (such as, for example, the lobe of the ear) is measured at two different spectral regions, so chosen that at one region the two pigments have the same absorption coefficient, while at the other the coefficients are different, then data may be obtained as to the relative abundance of the pigments, *i. e.* the degree of oxygenation of the blood may be estimated. It is, however, clear that this particular application merely forms one special case in a wide field of investigation. On the other hand, it should be emphasized that the models which are amenable to anything in the nature of a rigorous treatment are necessarily very much simplified descriptions of actual biological structures. Even so, it appears that the inherent complications of biological material will not seriously affect the results of our treatment as far as it goes, but care has been taken at each stage to indicate the nature of the simplifications which have been employed.

2. The Absorption of Light in Tissues.

A layer of tissues may be regarded for our purposes as consisting of a matrix of semi-transparent material through which passes a large number of blood-vessels. The first point we shall discuss is the manner in which the absorption of a ray of light traversing such a layer depends on the volume and distribution of the blood which it contains.

In this first approach to the problem of absorption we shall neglect scattering of light; we assume that the refractive index of the blood is the same as that of the surrounding tissues, and we restrict ourselves entirely to the straightforward problem of absorption. In addition, we take as a simplified model of our layer of tissues a plane parallel layer containing a large number of vessels, all circular in section, all of the same diameter, and all parallel to each other and to the surfaces of the layer. We shall endeavour later to generalize our results to a model bearing rather more resemblance to real tissues.

We denote the thickness of the tissue layer by t , and the radii of all the parallel vessels by a , and consider a section through the tissues

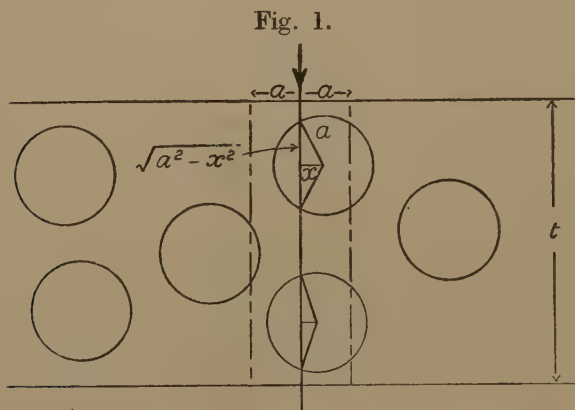
normal to the direction of the vessels. Let there be n vessels, distributed at random, per unit area of the section. We denote by κ_1 the absorption coefficient of the blood in the vessels, and by κ_2 that of the surrounding tissues.

Then, if a ray of intensity I_0 per unit area is incident normally on one surface of the tissue-layer, the transmitted intensity will be

$$I_0 e^{-\kappa_2(t-s_1-s_2\ldots)-\kappa_1(s_1+s_2+\ldots)}, \quad \text{. (1)}$$

where s_1, s_2 , etc., are the lengths intercepted on the ray by the vessels which it traverses. There are a large number of vessels, so that the absorption on the average normal ray may be found by taking the average value of the sum $s_1+s_2+\ldots$. Strictly speaking, the quantity required is the mean value of the transmitted intensity, but the argument advanced here will be incorrect only if some values of Σs occur which are very different from its mean value.

This mean value may be evaluated as follows:—A normal ray will intersect a vessel only if the centre of the vessel lies at a distance x , less than a , from the ray. (See fig. 1.) The length of such an intercept



will be $2\sqrt{a^2-x^2}$. Hence the vessels whose centres lie at distances between x and $x+dx$ on one side of the ray (their number being $n.t.dx$) will contribute a total intercepted length of $2.n.t\sqrt{a^2-x^2}.dx$, and the total contribution for all values of x on both sides of the ray will be

$$\int_0^a 4nt\sqrt{a^2-x^2}dx = \pi nta^2. \quad \text{. (2)}$$

Thus there is a zone, lying equally on each side of the ray and of total breadth equal to the diameter of a blood-vessel, within which lie the centres of all the vessels which contribute to the absorption of that ray.

The result of equation (2) may be generalized. Suppose we abandon the assumption that all the vessels have the same radius, and instead, assume that in each unit area of the cross-section there are $n(a).da$ vessels whose radii lie between a and $a+da$. The contributions from

each vessel-size are additive, and so the total average intercept on the transmitted ray made by blood-vessels now becomes

$$\pi t \int_0^A a^2 n(a) da, \quad (3)$$

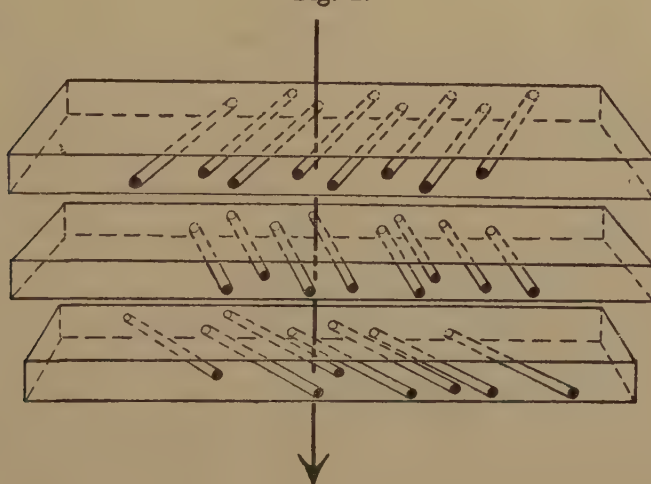
where A is the largest radius which occurs.

This result may be stated in a surprisingly simple form. We call the quantity B , defined by

$$B = \int_0^A \pi a^2 n(a) . da, \quad (4)$$

the *blood cross-section*, for it is the proportional area of the section occupied by blood-vessels. Hence equation (3) is simply equivalent to the statement that the average intercept made by blood-vessels on a normal ray is the product of the blood cross-section and the thickness of the layer.

Fig. 2.



It appears that a first approximation to an actual layer of tissue may be made by considering it as containing a number of sets of parallel blood-vessels, the vessels in all the sets being parallel to the same plane, but those in each set having a different orientation. (See fig. 2.) These sets are connected to each other by a comparatively small number of vertical vessels which we neglect in this approximate discussion.

If a ray passes normally through such a system of vessels the intercepts made by the several sets will be additive, so that the total intercept will be

$$t . (B_1 + B_2 + B_3 +),$$

where B_1, B_2 , etc., are the blood cross-sections of the individual sets of vessels evaluated for the neighbourhood of the normal ray. The total ΣB clearly represents the total proportionate volume occupied by vessels running in any direction at the site of the normal ray, so that *the total*

absorption of light, excluding scattering, depends only on the total volume of blood present and not on the manner of its distribution.

In so far as the conditions we have assumed are applicable to real tissues, this result is of some importance, since it suggests that deductions about oxygen concentrations made from observations of transmitted intensities are not likely to be falsified by the transference of blood from larger to smaller vessels, such as may occur in certain pathological conditions.

This result appears surprising at first sight, since it might have been argued that if a given quantity of blood were distributed among smaller vessels, presenting a greater surface in proportion to their volume, then the absorptive power of the same quantity of blood would have been increased. The fallacy of this argument may be seen as follows:—take the original case, illustrated in fig. 1, of a system of equal parallel vessels, and a section of the tissues normal to them. If the radius of each vessel is halved, their number per unit area must be increased four times if B is to be kept constant. Each vessel now gives half the intercept that it did before, and in addition, the zone on either side of the normal ray within which the vessel centres must lie if interception is to occur at all, is halved in width. The final value of the average total intercept is therefore unchanged.

3. Absorption in Some Other Structures.

The simplicity of the result for absorption of light in a structure consisting of a number of parallel tubes suggests that it might be worth while to investigate the absorption in structures of other types which may serve as models of animal or plant tissues, or in general for a situation in which one medium is dispersed in some manner throughout a second medium. There is one obvious restriction limiting the applicability of any results we may obtain in this way, namely, that of size. None of the results may be expected to be valid for structures in which the dimensions of the structural features are comparable with the wavelength of light, for we have neglected diffraction effects. In particular they will not apply to fine dispersions such as colloids, but there is a range of size above that of colloids where the results will be applicable. A diameter of $1-10\mu$ seems a reasonable lower limit of size for the structural dimensions.

The first case for investigation is clearly that of a system of spheres all of the same radius a , whose centres are distributed at random. The spheres whose centres lie at distances between x and $x+dx$ from the line of a given normal ray are $2\pi ntxdx$ in number, where n is the number per unit volume. Each of the spheres will contribute an intercept $2\sqrt{a^2-x^2}$, so that the total is

$$\int_0^a 4\pi ntx\sqrt{a^2-x^2}dx = \frac{4}{3}\pi nta^3 = B.t, \quad \dots \dots (5)$$

where B is the proportionate volume occupied by the spheres. This

result can obviously be generalized without change of form to the case when the radii of the spheres are not all equal, but when the statistical distribution of the radii follows some definite law.

More generally the result is true when we replace the spheres by solids of revolution, all of the same shape and size, and all with their axes parallel to the normal ray. Suppose the generating curve of a solid of revolution is defined by $y=F_1(x)$ for its upper portion, and $y=F_2(x)$ for the lower. Then, in this case, the average total intercept is

$$2\pi \int_0^\alpha nx\{F_1(x)-F_2(x)\}dx,$$

where α is the horizontal semi-axis of the generating curve. But the volume of each of the solids is

$$2\pi \int_0^\alpha x\{F_1(x)-F_2(x)\}dx,$$

and so again the average total intercept is $B.t.$

It is hardly to be expected that in a natural structure the axes of all the absorbing elements will be parallel. We can illustrate the effect of randomizing the directions of the axes by considering the case where the absorbing structure consists of a series of thin circular discs, whose axes are orientated at random in space, and whose centres are also distributed at random in space.

Let the radius of each disc be r and its thickness k . Let the inclination of the axis of a disc to the direction of the normal ray be θ . Then, if the distribution of directions of these axes be random, the proportion lying in the range θ to $\theta+d\theta$ is $\sin \theta d\theta$. Then the number of discs whose centres lie at distances between x and $x+dx$ from the normal ray is $2\pi x dx . n . t$, where n is the number of discs per unit volume. A disc at angle θ contributes $k \sec \theta$ to the intercept, so that the total average intercept is given by

$$\int_{x=0}^r \int_{\theta=0}^{\cos^{-1} x/r} 2\pi x dx n . t . k . \tan \theta d\theta = \frac{\pi}{2} r^2 n t k = \frac{1}{2} t . B, \quad (6)$$

the limits of this integral being dictated by the fact that values of θ greater than $\cos^{-1} (x/r)$ make no contribution to the intercept. (Fig. 3.)

The randomization of the direction of the axes for this particular type of solid of revolution reduces the average intercept to one-half. It does not appear that results of equal simplicity apply to other solids of revolution. However, in all the cases quoted, it remains true that as far as pure absorption is concerned the diminution in intensity depends only on B , *i. e.*, the total amount of the dispersed medium which is present, and not on the manner in which it is distributed among absorbing elements of a given type.

4. *The Scattering of Light.*

Two mechanisms seem possible by which light may be scattered in media of the type which we have been considering. The first is by

reflection from the surfaces of the various elements (tubes, spheres, discs, etc.) which compose the structure. The second is by ordinary scattering in the continuous mass of the media, which, in biological structures containing large molecules, seems likely to be very considerable.

A. Surface Scattering.

Some idea of the operation of the first effect may be obtained by considering the case of diffuse scattering of a collimated beam incident normally on a circular tube of radius a . Suppose the incident beam is

Fig. 3.

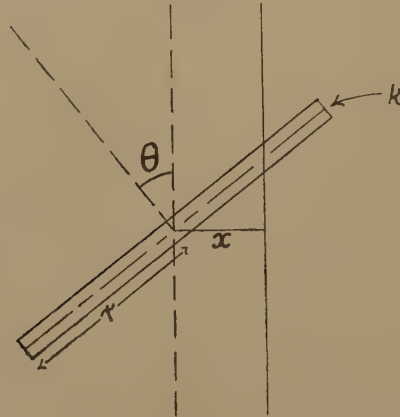
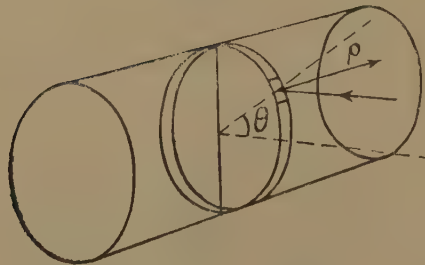


Fig. 4.



of intensity I per unit area, and we consider scattering from a short length, ds , of the tube-surface. Then the incident flux on an element of area $a \cdot d\theta \cdot ds$ at the point θ (fig. 4) is $I \cdot a \cdot d\theta \cdot ds \cdot \cos \theta$. We suppose that a small fraction η of this incident flux is diffusely reflected. Marking the directions on a unit sphere (fig. 5), we find that the diffused intensity per unit solid angle in a direction defined by the angles ϕ and χ is proportional to

$$I a \cdot ds \cdot d\theta \cdot \eta \cos \theta \cos \rho.$$

By the formulæ of spherical trigonometry

$$\cos \rho = \sin \theta \sin \phi + \cos \theta \cos \phi \cos \chi. \quad . \quad . \quad . \quad . \quad (7)$$

Now at a distance from the tube, which is large compared with its radius, the length ds of the tube will behave like a point source for which all the contributions to the flux in direction (ϕ, χ) may be summed. Contributions to this sum will be received from all parts of the illuminated half of the element which lie within a right angle from the given direction.

Thus the upper limit of the integral is given by $\theta = \frac{\pi}{2}$ and the lower by $\theta = \theta_0$, where $\tan \theta_0 = \cot \phi \cos \chi$.

Fig. 5.

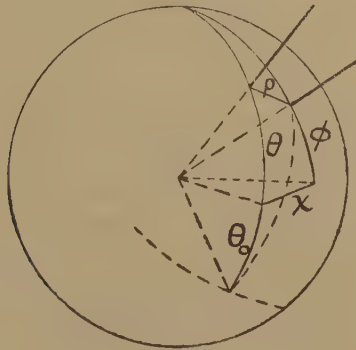
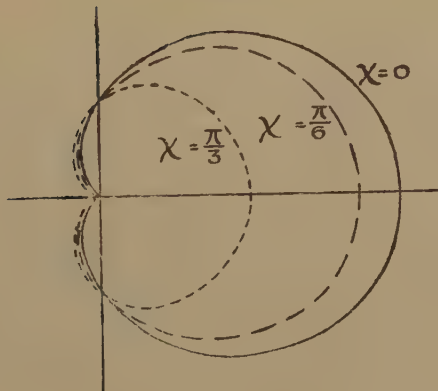


Fig. 6.



The trigonometrical part of the integral gives

$$\frac{\sin \phi}{2} \cos^2 \theta_0 + \frac{\cos \phi \cos \chi}{2} \chi \left(\frac{\pi}{2} + \theta_0 + \sin \theta_0 \cos \theta_0 \right), \quad \dots \quad (8)$$

when $\chi = 0$, $\theta_0 = \frac{\pi}{2} - \phi$, and (8) becomes

$$\frac{\sin \phi}{2} + \frac{\pi - \phi}{2} \cos \phi. \quad \dots \quad (9)$$

Polar plots of reflected intensity are shown in fig. 6.

It will be noted that the expression (9) applies to values between θ and π and that it is not an even function of ϕ . The values for negative values of ϕ are to be obtained by reflection of the upper half of the curve in the horizontal axis. The reason for this is that when negative values of ϕ are in question, or values greater than π , the limits of the integral have to be changed to a range running from $-\frac{\pi}{2}$ to $+\theta_0$. It should also be noted that the reflected intensity is proportional to a and not to B , *i. e.*, this effect increases as a given volume of blood is distributed among smaller vessels.

As a first approximation, therefore, effects of this kind may be taken into account by postulating an additional small scattering coefficient which produces isotropic scattering, and which increases as the dimensions of the structural features are reduced.

A word may also be said here about the effects of refraction.

A system of spheres or tubes will act as a large number of weak lenses which may be roughly represented (pending a more detailed treatment) as producing a small scatter, corresponding to the deviation of rays passing through them, while the remainder passes on with slightly diminished intensity. The scattering corresponding to this effect is of course not isotropic, which renders its treatment difficult. However, we may, to a first approximation, take account of all these effects by including a suitable scattering coefficient.

B. *Volume Scattering.*

When we consider scattering by the media themselves we find ourselves on much more certain ground. As a ray passes through a medium two contributions are abstracted from it. One of these is absorbed, the average total path length being, as we have seen, simply related to B . The second contribution is re-emitted isotropically as scattered radiation. Now consider any element of path. The absorbed radiation is proportional to the length of the element and to the absorption coefficient of the medium. The scattered radiation is proportional to the path-length and the scattering coefficient. Thus, if the two coefficients are constant, then there is a constant ratio between the absorption and scattering on any element of path.

In order to deal with the problem mathematically we replace the actual tissues by a theoretical homogeneous medium, for which the absorption and scattering coefficients are constant and have values which are the same as the average scattering and absorption coefficients of the actual structure.

If B is the blood cross-section, and κ_1 and κ_2 the absorption coefficients of the blood and the surrounding tissues respectively, then the required average absorption coefficient is given by

$$\kappa = (1 - B)\kappa_2 + B\kappa_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

But since, for constant coefficients, the scattered light produced in a

given path-length is in a constant ratio to the absorption, then the scattering coefficient of the "smoothed" medium has a value given by a similar mathematical form

$$\sigma = (1-B)\sigma_2 + B\sigma_1, \quad \dots \quad (11)$$

To this must be added a further term, which may be small, to account for other scattering mechanisms such as diffuse reflection, refraction, etc., giving

$$\sigma = (1-B)\sigma_2 + B\sigma_1 + k. \quad \dots \quad (12)$$

Now the energy abstracted from a short length of path is proportional to the sum of κ and σ , so that, as far as directly transmitted light is concerned, the tissue-layer behaves as if it had an absorption coefficient of $\sigma + \kappa$. If, therefore, the additional term, k , in (12) is small enough to be negligible, the conditions in the tissue-layer depend only on B , and the absorption and scattering coefficients of the blood and tissues, *i. e.* in these circumstances it remains true that the transmitted intensity depends only on B and not on the manner of the distribution of the given quantity of blood.

The general equation satisfied by the rate of flow of radiant energy I per unit area per unit solid angle in a given direction at a point in the tissues, is defined by the equation

$$\frac{dI}{ds} = -(\kappa + \sigma)I + \frac{\sigma}{4\pi} \int I d\omega, \quad \dots \quad (13)$$

where the integral is to be taken over a unit sphere, ds is the element of path in the given direction, and κ and σ are defined by equations (10) and (12) above. The solution of equation (13) must be adjusted to comply with the given boundary conditions.

The case in which we are particularly interested is that of a collimated beam of radiation incident normally on a layer of tissues. This may be regarded as the limiting case of a problem in which the incident radiation is all confined in direction within a small solid angle ω described about the normal. Thus, if I' is the incident intensity, per unit area per unit solid angle, then I' is supposed to be zero except within the solid angle ω . To take the limit we suppose that $I' \cdot \omega =$ a constant I_0 but that ω tends to zero. In the limit the flux per unit area of the collimated beam will be I_0 .

Now consider equation (13) and the energy flow normal to the surface. If the equation is multiplied throughout by ω and the limit taken, both $I \cdot \omega$ and $\frac{dI}{ds} \cdot \omega$ will tend to a definite limit. On the other hand, the

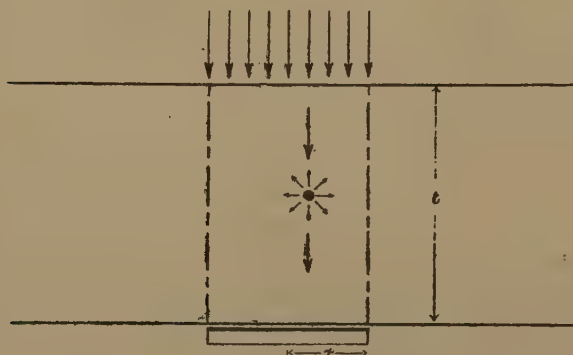
integral term will tend to zero, because not only is there a factor ω outside the integral itself, but the integrand is non-zero only in the solid angle ω . Hence it follows that if we have to consider a perfectly collimated beam, the scattering *exactly* in the direction of the incident beam will be negligible, and the only effect of scattering will be to

produce an apparent increase in the absorption coefficient from κ to $\sigma + \kappa$.

However, in a practical case we have to concern ourselves not only with light scattered exactly in the direction of the incident beam, but with light scattered within some finite angle. Consider, for example, the situation depicted in fig. 7, where a collimated beam of light, circular in section and of radius r , and having an intensity I_0 per unit area, is incident normally on a tissue-layer of thickness t . Let the transmitted light be received on an equal circular area, *e.g.* a photo-cell of some kind. Then light scattered from an element of volume such as that indicated in the diagram can reach the receiver for a finite range of the scattering direction.

We shall endeavour to estimate the amount of this scattered light in the case indicated. Now if we trace out in detail the fate of the energy incident at a certain point on the surface, we find that for each element

Fig. 7.



of path traversed there is a loss of energy, the proportion lost varying with ϕ , where

$$\phi = \sigma + \kappa. \quad (14)$$

The proportion of this depending on σ is re-radiated and forms a new ray which in its turn undergoes absorption and scattering. In order to reduce the calculation to manageable proportions we shall calculate what is in effect a lower limit to the total light scattered on to the receiver. We shall sum the radiant energy which falls on the receiver after being scattered *once* only, and compare this total with the total direct radiation, *i.e.* with the radiant energy falling on the receiver which has been scattered *zero* times. Double and triple scattering and scattering of higher multiplicity will produce small further contributions, but the ratio of each contribution to the preceding one will be of the same order as the ratio of the once scattered light to the direct light.

From what has been said it will be seen that the intensity of the direct beam at a depth s below the surface will be given by

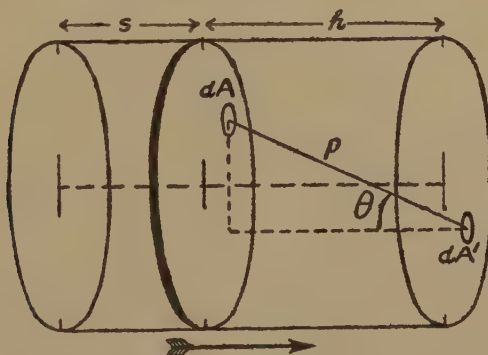
$$I = I_0 e^{-\phi s}. \quad (15)$$

Then a disc of radius r and thickness ds at a depth s below the surface (fig. 8) becomes a source of (once scattered) energy emitting $I_0 e^{-\phi s} \cdot \sigma$ per unit volume. Thus a surface element dA will transmit to a surface element dA' of the receiver a flux

$$I_0 e^{-\phi s} \cdot \sigma ds dA \cdot \frac{dA' \cos \theta}{4\pi \rho^2} e^{-\phi \rho},$$

where ρ is the distance between dA and dA' and θ the angle between the line joining them and the direction of the normal ray. The factor $\frac{dA' \cos \theta}{4\pi \rho^2}$ is the element of solid angle which dA' subtends at dA , and $e^{-\phi \rho}$ is the fraction of the scattered intensity which finally arrives at dA' . It should be noted that the expression includes a factor $\cos \theta$ and not $\cos^2 \theta$, the reason being that, although dA presents to dA' a projected area of only $dA \cos \theta$, this reduction is offset by the fact that the line

Fig. 8.



joining the two elements of area has a length $ds \cdot \sec \theta$ within the disc of normal thickness ds .

The total scattered radiation which therefore arrives on the receiver is given by

$$I_0 \frac{\sigma}{4\pi} \int_{s=0}^t e^{-\phi s} \iint dA \cdot \frac{dA' \cos \theta}{\rho^2} e^{-\phi \rho} ds, \quad \dots \quad (16)$$

where the integrals of area are to be taken over circles of radius r .

The integral (16) is hardly susceptible of exact evaluation and so, in the portion

$$\iint dA \cdot \frac{dA' \cos \theta}{\rho^2} e^{-\phi \rho}, \quad \dots \quad (17)$$

we approximate by replacing ρ by its mean value $\bar{\rho}$. This may be evaluated as follows:—Consider two coaxial discs of equal radii r , their centres being distant h apart. Take a point P in one disc at a distance c from its centre. The mean distance of P from all points in the second

disc is the distance from P to its centre (of gravity), namely $\sqrt{c^2+h^2}$. The required value $\bar{\rho}$ is thus given by

$$\bar{\rho} = \frac{1}{\pi r^2} \int_0^r 2\pi c \, dc \sqrt{c^2+h^2} = \frac{2}{3} r \chi, \quad (18)$$

where
$$\chi = \left(1 + \frac{h^2}{r^2}\right)^{3/2} - \frac{h^3}{r^3}. \quad (19)$$

If in the integral (17) we write $h=t-s$, then $\rho \cos \theta = h$, and the integrand becomes $he^{-\phi\rho/\rho^3}$, which, on replacing ρ by $\bar{\rho}$ gives

$$\frac{27}{8r^3} \frac{he^{-\frac{1}{3}\phi\chi}}{\chi^3}. \quad (20)$$

With these changes the integral (16) becomes

$$I_0 \sigma \pi r \frac{27}{32} e^{-\phi t} \int_0^t e^{\phi h} \frac{h}{\chi^3} e^{-\frac{1}{3}\phi\chi} dh. \quad (21)$$

Let Θ denote the ratio of the scattered light to the transmitted light. The latter is $I_0 \pi r^2 e^{-\phi t}$, so that

$$\Theta = \frac{\sigma}{r} \frac{27}{32} \int_0^t e^{\phi h} \frac{h}{\chi^3} e^{-\frac{1}{3}\phi\chi} dh. \quad (22)$$

Write $t' = \phi t$, $r' = \phi r$ and $h' = \phi h$, so that

$$\chi = \left(1 + \frac{h'^2}{r'^2}\right)^{3/2} - \frac{h'^3}{r'^3}.$$

We calculate the integral

$$S = \frac{1}{r'} \frac{27}{32} \int_0^{t'} \frac{h'}{\chi^3} e^{h'} e^{-\frac{1}{3}h'} dh'. \quad (23)$$

numerically, and then

$$\Theta = \frac{\sigma}{\sigma + \kappa} S. \quad (24)$$

The calculated values are given in Table I. and illustrated in fig. 9.

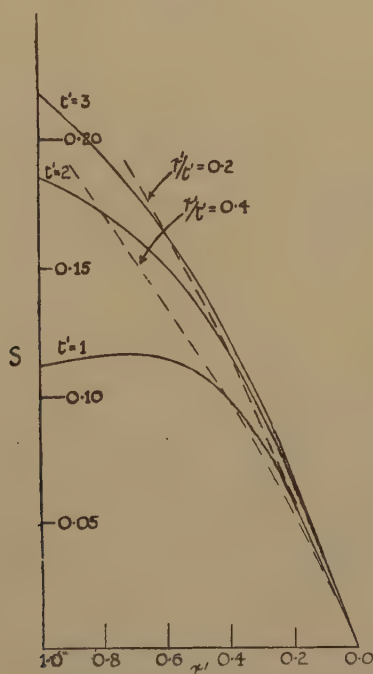
TABLE I.

t'	Values of S.					Percentage of incident light transmitted = $100 e^{-t'}$.
	$r'=1.0$	0.75	0.5	0.2	0.1	
1.0	0.112	0.117	0.110	0.057	0.036	36.8
2.0	0.186	0.168	0.137	0.062	0.036	13.5
3.0	0.218	0.188	0.146	0.062	0.036	5.0

In fig. 9, the dotted lines show the values of S for constant values of r'/t' .

Equation (24) shows that the proportion of scattered light depends not only on S but on the ratio of σ to κ . Now the usual methods of determining absorption coefficients spectrophotometrically clearly determine, not κ , but an approximation to $\sigma + \kappa$ whose closeness depends on the ratio of the diameter of the receiver to the thickness of the layer of tissue or fluid which is the subject of the experiment. The author is not aware of any determinations of σ itself—determinations which would necessitate simultaneous observations both in the direction of the incident beam and perpendicular to it. We have, therefore, no definite data on which to rely in estimating the value of the fraction $\frac{\sigma}{\sigma + \kappa}$, but workers

Fig. 9.



experienced in medical and biological spectrophotometry with whom this point has been discussed seem to consider that in the case of many biological media σ may well be as large as κ . If this is so, then $\frac{\sigma}{\sigma + \kappa}$ must be of the order of $1/2$, so that from fig. 9 we see that, with a layer of tissue of optical thickness 3, and a receiver diameter of 2, over 10 per cent. of the recorded intensity may be due to scattered light. In some instruments which have been described the "aperture ratio" has been even greater than this. The figure also shows that S can be reduced to less than 0.05 by taking r' less than about 0.2,

It is perhaps worth while calling attention to one case in which the consideration of σ separately from κ may provide an explanation of an observational anomaly. The observed absorption curves of intact red blood-cells and of hæmolyzed solutions of hæmoglobin differ considerably (Keilin and Hartree, 1941). Both these curves show the variation of $\sigma + \kappa$ with wave-length and not of κ alone, and it may be that the differences are to be explained as differences in σ rather than as differences in κ itself.

5. *Acknowledgments.*

The present work was undertaken as part of a general investigation on the applicability of blood oxygen meters carried out by Dr. K. Mendelssohn and the author at the Clarendon Laboratory, Oxford. Particular attention was paid to the question of distribution in size of blood-vessels in response to an enquiry by Professor John Beattie of the Royal College of Surgeons, to whom the author is indebted for helpful discussions on a number of points, more especially those of physiological interest. The author wishes to acknowledge his indebtedness to Dr. K. Mendelssohn, who pointed out the more general applicability of the problem, for his advice and encouragement, and to Mr. E. M. Jope of the Department of Biochemistry, Radcliffe Infirmary, Oxford, for discussions of spectrophotometric points.

6. *Summary.*

The absorption and scattering of light in media of two constituents, in which the second constituent is distributed throughout the first in the form of a series of vessels having some common geometrical characteristic, has been considered. The particular application of the results to the transmission of light through living tissues has been discussed, and it has been shown that, on certain plausible assumptions, the absorption of light in a layer of blood-bearing tissues depends only on the total volume of blood present and not on the manner of its distribution as between larger and smaller vessels. The influence of scattered light has been considered, and it has been shown that its proportionate effect in the total energy transmitted is reduced when the diameter of the beam is diminished.

References.

- Goldie, J. Sci. Inst. xix. p. 23 (1942).
Keilin and Hartree, 'Nature,' cxlviii. p. 75 (1941).
Millikan, J. Sci. Inst. xiii. p. 434 (1942).
Squire, Clin. Sci. iv. p. 3 (1940).

XXXVIII. *A new Approach to the Dynamics of Systems with Gyroscopic Coupling Terms.*

By A. BLOCH, Dr.-Ing., M.Sc.*

(Communication from the Staff of the Research Laboratories of
The General Electric Company, Limited, England.)

[Received December 3, 1943.]

Summary.

Passive electrical networks and "ordinary" mechanical systems have in common that mutual impedances (admittances) are symmetrical: $z_{ik} = z_{ki}$. This is basic for the use of the electrical analogies.

In mechanical systems with gyroscopic couplings this symmetry gives place—in part—to a skew symmetry: and, as far as the author is aware, such systems have for this reason not been treated by means of electrical analogies.

In the present paper the concept of an ideal gyroscopic coupler is introduced, and it is shown that a mechanical system which is driven from the secondary side of such a coupler "looks" from the primary side like the reciprocal (dual) counterpart of itself. We may therefore omit the gyroscopic coupler entirely and connect this "transformed" system directly to the system connected to the primary side. The resulting system, which is now free from gyroscopic couplings, can then be treated by means of either of the electrical analogies. The electrical representation of this system can also be obtained by direct translation from the original system, say by translating the primary part according to the direct analogy, the secondary part according to the rules of the inverse analogy.

A number of applications are given to systems containing one or more gyrostatic couplers.

The structure and behaviour of systems to which the method is applicable are discussed. It is shown that, when the method becomes inapplicable, it leads to the finding of a class of mechanical systems which possess asymmetrical impedance properties, corresponding to electrical systems containing valves. A simple example for such a system is given, and a representative electrical circuit is developed which contains only passive elements.

* Communicated by C. C. Paterson, F.R.S.

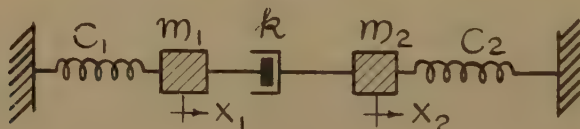
CONTENTS.

	Page
1. Introduction	316
2. The Ideal Gyroscopic Coupler	317
3. Reciprocal Systems	320
4. Real Gyroscopic Couplers: Nutations of a Free Top.....	323
5. The Gyroscopic Pendulum: Two Modes of Precession	324
6. The Use of the Electrical Analogy	325
7. The Gyroscopic Ship's Stabilizer in Terms of the Electric Analogy	328
8. Further Applications:	
(a) Pseudo-regular Precession.....	329
(b) Coupling due to Coriolis Forces	329
9. Systems with more than one Gyroscopic Coupler	331
10. The Structure of Systems to which the Present Method is applicable: Remarks concerning the Behaviour of such Systems	332
11. Systems with Asymmetrical Impedance Properties	333

1. Introduction.

"ORDINARY" mechanical oscillatory systems are characterized by the existence of a well-known symmetry relation between the coefficients of the governing set of differential equations. As an example we shall

Fig. 1.



write down the equations for the free oscillations of two mass points, each of which is controlled by a spring and which are coupled together by means of a viscous resistance (fig. 1):

$$\left. \begin{aligned} m_1 \ddot{x}_1 + c_1 x_1 + k(\dot{x}_1 - \dot{x}_2) &= 0, \\ m_2 \ddot{x}_2 + c_2 x_2 + k(\dot{x}_2 - \dot{x}_1) &= 0. \end{aligned} \right\} \quad \dots \quad (1)$$

The coupling term entering the first equation is $-k\dot{x}_2$, the term entering the second equation is $-k\dot{x}_1$; and we note that the coefficient with which the "intruding" velocity is multiplied is of the same magnitude and sign in both equations. In a more general case we might have met couplings arising from mutual masses or elasticities, and we would have found the same kind of symmetry concerning the coefficients with which the "intruding" accelerations or displacements are multiplied; in the following we are, however, specially interested in couplings arising from velocity terms.

Our knowledge of the behaviour of mechanical systems possessing this symmetrical type of coupling has greatly benefited by the intense study devoted to the problems of electrical networks. Passive electrical

networks are governed by differential equations of an exactly analogous type, and the use of electro-mechanical analogies for the solution of mechanical problems is therefore steadily increasing. The symmetry to which attention is here drawn is well known in such networks in the forms $z_{ik}=z_{ki}$ (the mutual impedance between two meshes is equal) or $y_{ik}=y_{ki}$ (the mutual admittance between two junctions is the same in both directions).

It is obvious from the foregoing that these electro-mechanical analogies are of no immediate use for the study of mechanical systems with gyroscopic coupling terms, *i. e.* terms arising from the velocities and in which the symmetry mentioned has been replaced by a skew symmetry. For this reason it is thought to be of interest that in many cases (amongst which most practically important cases are contained) it is possible to transform by elementary considerations the mechanical system into another mechanical system which contains only "ordinary" couplings, and which can, therefore, be studied by means of the electrical analogies.

That this possibility exists is probably immediately obvious to those who have studied the interaction of electrical and mechanical networks due to electrodynamic forces*. This type of interaction is characterized mathematically by gyroscopic terms, and the fact that we can study the combined network in the form of a purely electrical or purely mechanical network of the "ordinary" type points at once to the possibility of a similar transformation in the purely mechanical case. The treatment in the present note, while mentioning this analogy, does not, however, rely on it for the derivation of the desired transformation. It discusses it first in purely mechanical terms, and gains thereby a better understanding of the behaviour of such systems than could be attained by a forward and backward translation into the electrical analogy.

2. The Ideal Gyroscopic Coupler.

We begin by discussing an ideal gyroscopic coupler such as illustrated in fig. 2. This figure shows a gyrostator (flywheel) in its conventional mounting. If the flywheel is rotating, then movements around axis 1 are coupled to movements around axis 2. We call the mechanism shown an ideal coupler because we suppose the following special conditions to be fulfilled :

- (a) The gimbals are frictionless and massless.
- (b) The moments of inertia Θ_{10} and Θ_{20} of the flywheel around axes 1 and 2 are negligible, ideally zero.

The last condition seems to contradict the requirement that there should be a finite angular momentum $B=\Theta_3\omega_3$ around axis 3, because

* See the author's paper "Electro-Mechanical Analogies and their use for the Analysis of Mechanical and Electro-Mechanical Systems." (Accepted for publication in the 'Journal of the Institution of Electrical Engineers.')

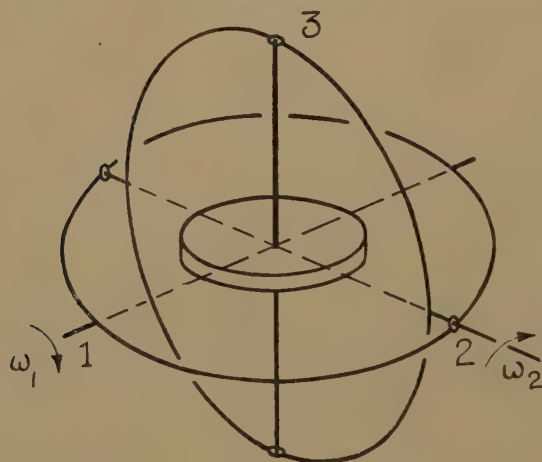
$\Theta_3 \leq \Theta_{10} + \Theta_{20}$. However, in this ideal case we are allowed to stipulate a value of ω_3 sufficiently high, so as to give the product $\Theta_3 \omega_3$ a finite value, even if Θ_3 approaches zero.

Another assumption made in the following discussion is that the angles ϕ_1 and ϕ_2 formed by the axis 3 with the normal to the plane containing axes 1 and 2 remain small during the operation of the coupler, so that $\sin \phi \approx \phi$ and $\cos \phi \approx 1$. This assumption is frequently permissible in investigations of this type.

Let us now "operate" this coupler. Suppose we impose an angular velocity ω_1 about axis 1. Then it follows from the law of angular momentum that we require a moment

$$M_2 = B\omega_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Fig. 2.



acting on axis 2 in order to prevent this axis from turning. Should ω_1 vary in time as $e^{j\Omega t}$, then M_2 will vary in the same way, and we may write the last relation as a relation between the time vectors $\hat{\omega}_1$ and \hat{M}_2 :

$$\hat{M}_2 = B\hat{\omega}_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Under the assumptions made, the moment M_1 required to bring about this oscillating movement around axis 1 is zero.

Let now the constraint on axis 2 be insufficient to prevent it from turning, so that it attains an angular velocity ω_2 . Such an angular velocity requires of course a moment M_1 to be imparted to axis 1,

$$M_1 = \pm B\omega_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

if we wish this axis to move in the same way as before.

We have written the last equation with a \pm sign, as a determination of the correct sign, though simple, need not trouble us at present.

they are valid for all values of Ω from $-\infty$ to $+\infty$, or, on account of Fourier's Theorem, that they are valid not only for movements following a sinusoidal time law, but that they are valid for any time law.

This last result could also have been obtained from first principles, *e. g.* if $\phi_2 = M_2 \Delta_2$, then

$$\omega_2 = \frac{d\phi_2}{dt} = \Delta_2 \frac{dM_2}{dt} = \Delta_2 B \frac{d\omega_1}{dt} \quad \text{and} \quad M_1 = B\omega_2 = \Delta_2 B^2 \frac{d\omega_1}{dt}, \quad (11)$$

which last equation allows the same interpretation as equation (9). However, it seems quite attractive to carry out the present investigation by making the use of "impedance" concepts only*.

3. *Reciprocal Systems.*

A "translation" of the type here described is familiar to the electrical engineer. According to the Principle of Duality it is possible to construct to every electrical network a counterpart in which the rôles of voltage and current are interchanged; the relation between corresponding parts of two such dual networks is just the one given in equation (7) (replacing the mechanical impedances by electrical impedances). It is on account of this last relation that the process of transforming a network into its dual counterpart is also shortly called "reciprocation."

Reciprocation is a useful mathematical device which allows us to utilize the solution found for one type of network at the same time—merely by interchanging the words current and voltage—for another network. The principle of duality holds, of course, also in the field of Mechanics, and from the discussion in the preceding section it is seen, that the ideal gyroscopic coupler is a device which would be capable of carrying this process physically into effect.

The rules for the determination of the reciprocal network are best derived by a direct application of the Principle of Duality. However, it is interesting to note that relation (7) also contains these rules implicitly.

Let us, for instance, assume that the system connected to axis 2 of the coupler consists of two elements which are connected "co-resistively"†, *i. e.* that we have

$$\hat{z}_2 = \hat{z}_a + \hat{z}_b. \quad (12)$$

Then the admittance seen on the primary side of the coupler is

$$\hat{y}_1 = \frac{\hat{z}_2}{B^2} = \frac{\hat{z}_a}{B^2} + \frac{\hat{z}_b}{B^2} = \hat{y}_{1a} + \hat{y}_{1b}. \quad (13)$$

* When non-linear impedance elements have to be dealt with, it is, of course, simpler to go back to the differential equations in the way just instanced, *i. e.* to the "Principle of Duality" testified by these equations.

† As the author has already pointed out in the paper quoted above, the use of the ambiguous terms "parallel-" and "series-connection" is best avoided in the description of mechanical systems. Alternative designations suggested are "co-resistive" (when the forces add up) and "co-yielding" (when the displacements add up).

where \hat{y}_{1a} and y_{1b} are the admittance of the "transforms" of z_a and z_b individually. This means that the "transforms" of two co-resistively connected elements have to be connected co-yieldingly.

Conversely, if two elements on the secondary side of the coupler are connected co-yieldingly, *i. e.* if

$$\hat{y}_2 = \hat{y}_a + \hat{y}_b, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

then we have on the primary side,

$$\hat{z}'_1 = B^2 \hat{y}_2 = B^2 \hat{y}_a + B^2 \hat{y}_b = \hat{z}'_{1a} + \hat{z}'_{1b}. \quad (15)$$

This means that the transforms have to be connected co-resistively.

The two rules just given are sufficient to find the reciprocal system in all those simple cases here discussed. For the sake of completeness the procedure to be followed in more complicated cases will be outlined. In such cases the electrical engineer uses a method which Guillemin* ascribes to Brune, which, however, goes back in its topological aspects to certain investigations carried out at the end of the last century in connection with the colouring of maps.

We shall describe this method as if we were proceeding to draw the reciprocal circuit diagram on top of the given diagram. We have then to select one point in the interior of each mesh of the given diagram: these points form the junction points of the new diagram. The branches of the new diagram are obtained by interconnecting all "neighbouring" junction points (*i. e.* junction points situated in neighbouring meshes). The admittance of such a new branch is $\hat{y}_{\text{new}} = \hat{z}_{\text{old}} / R_0^2$, where \hat{z}_{old} denotes the impedance of the old branch crossed by a new branch, and where R_0 , of the dimension of a resistance, denotes the constant of reciprocation. If \hat{z}_{old} consists of several impedances in series we draw a corresponding number of new branches in parallel; if the old branch contains a source of e.m.f., E , the new branch, has to contain a source of current $I = E/R_0$, and *vice versa*. The rule for the reciprocation of the impedance values includes at the same time the instruction to change the nature of the impedance elements: an inductance changes into a capacitance and *vice versa*.

To be able to apply these topological concepts to dynamical systems we have to draw diagrams which indicate the "flow of force" through such systems in a way first shown by Hähnle†.

* E. Guillemin, 'Communication Networks,' vol. ii. p. 285. The method is limited to "planar" networks, *i.e.* networks whose circuit diagrams can be plotted without crossings between branches. "Non-planar" networks have no topological duals (Whitney, Proc. Nat. Acad. Sci. vol. xvii. p. 125 (1931)). However, it is always possible, by the aid of ideal transformers, to construct a network which has an admittance matrix numerically equal to a given impedance matrix (and *vice versa*) (Cauer, *Electr. Nachrichten Technik*, vol. xi. p. 157 (1932)).

† Hähnle, *Wissenschaftliche Veröffentlichungen a.d. Siemens-Konzern*, vol. xi. no. 1 (1932).

It is easily seen that a massless spring or a massless "dashpot" possess two terminals, and any force applied to one of their terminals must appear with undiminished magnitude at their second terminal. We may therefore liken the application of force to such elements to the flow of some indestructible agent through these elements; the analogy to the flow of current is obvious. Masses appear to form an exception to this scheme. However, we need only link each mass to "ground" by means of some fictitious linkage and to ascribe the mass reaction to the properties of this fictitious linkage in order to have this remedied. True enough, the grounded terminal of the mass is not accessible, and we cannot observe

Fig. 3.

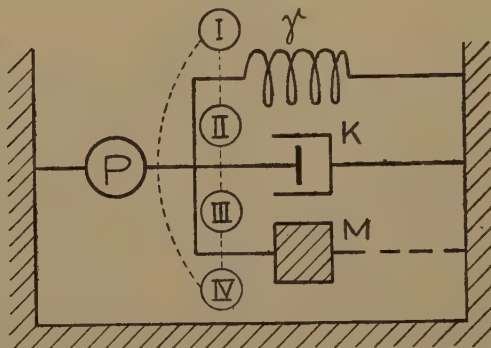
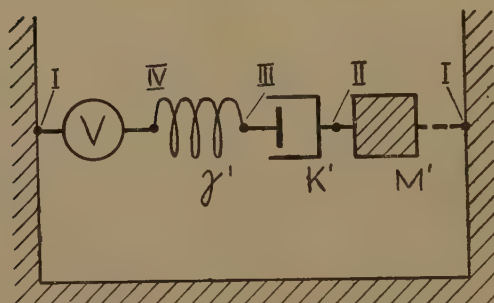


Fig. 4.



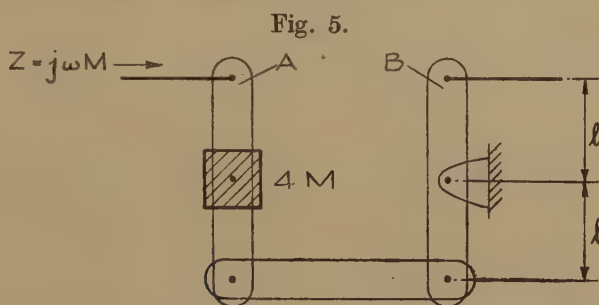
$$V = \frac{P}{K_0}, \quad \gamma' = \frac{K_0^2}{M}, \quad K' = \frac{K_0^2}{K}, \quad M' = \frac{K_0^2}{\gamma}.$$

the flow of force when it leaves this terminal. However, the "driver" of the mass (the source of the force entering the accessible terminal of the mass) needs a "foothold" on the ground, and at this foothold we can observe the entry of our flowing agent.

Thus the methods of the electrical engineer become available for the determination of reciprocal mechanical systems. An example is given in figs. 3 and 4. Fig. 3 illustrates the flow of a force P through a co-resistive combination of spring, dash-pot and mass. I, II, III and IV denote the new junction points. The reciprocated circuit is shown in fig. 4.

Attention has to be drawn to one difficulty concerning the "building" of such reciprocal mechanical systems, a difficulty already discussed and solved by Hähnle. In all those cases where the original system contains two meshes with a common elasticity, the dual system contains two junction points which are to be connected by a "massy" admittance. As the fictitious second terminals of all masses are permanently grounded, we have to resort to an artifice to fulfil this condition. It consists of the assembly of massless levers, shown in fig. 5, which is loaded with a mass $4M$ if the impedance offered against a relative movement of the free ends of the levers A and B is to be $j\omega M$.

It is the necessity of occasionally introducing such lever arrangements which makes the direct reasoning in mechanical terms liable to be more cumbersome than the reasoning in electrical terms, at least as far as the drawing of the necessary diagram is concerned.



4. Real Gyroscopic Couplers: Nutation of a Free Top.

Having discussed the behaviour of the ideal gyroscopic coupler it is now not difficult to extend the investigation to real gyroscopic couplers. These differ from our ideal ones at least in the finite magnitude of the moments of inertia Θ_1 and Θ_2 of gimbals and flywheel around axes 1 and 2.

Now we can consider these moments of inertia simply as permanent "loadings" applied to these axes. If no further loading is present, an oscillating moment M_1 applied to axis 1 has to overcome the impedance caused by Θ_1 as well as the impedance of the "transform" of Θ_2 , i. e. the impedance of an elasticity of stiffness

$$\gamma'_2 = \frac{B^2}{\Theta_2} \quad (16)$$

As longitudinal oscillations are more convenient for a diagrammatic representation than torsional oscillations, we shall represent our mechanical system in terms of equivalent "longitudinal" magnitudes and obtain the diagram fig. 6. We notice at once that the system is capable of free oscillations, the period of which is given by

$$T = 2\pi \sqrt{\frac{\Theta_1}{\gamma'_2}} = 2\pi \sqrt{\frac{\Theta_1 \Theta_2}{B^2}} \quad (17)$$

This period is, of course, nothing but the period of nutation of a gyroscope which is supported at its centre of gravity and not subject to other external forces.

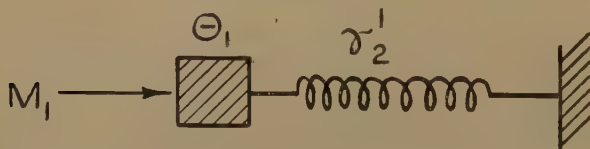
For $\Theta_1 = \Theta_2 = \Theta$ (17) simplifies to

$$T = 2\pi \frac{\Theta}{B} = 2\pi \frac{\Theta}{\Theta_3 \omega_3} = T_3 \cdot \frac{\Theta}{\Theta_3}, \quad \dots \quad (18)$$

where T_3 is the time required for one revolution of the gyrostatt.

We shall discuss now the ratio of the amplitudes ϕ_1 and ϕ_2 of these free oscillations around axes 1 and 2 as an example of the re-translation of results from the equivalent system shown in fig. 4 to our real system. The displacement ϕ_1 of the mass Θ_1 is a direct representation of the angular displacement ϕ_1 of axis 1; the force $M'_1 = \gamma'_2 \phi_1$ exerted by the spring on this mass represents the action of the velocity $\omega_2 = \dot{\phi}_2$ transmitted through the coupler: $M'_1 = B \dot{\phi}_2 = \gamma'_2 \phi_1$. Therefore at the instant where ϕ_1 reaches a maximum value there is also a maximum value of $\dot{\phi}_2$. If ϕ_1 follows a sine law (in time), $\dot{\phi}_2$ also follows a sine law, i. e. ϕ_2 follows a cosine law. The axis of the gyrostatt describes, then, a cone of elliptical or circular cross-section.

Fig. 6.



As the maximum values of ϕ_2 and $\dot{\phi}_2$ are connected by the relation

$$\dot{\phi}_{2\max} = \frac{2\pi}{T} \phi_{2\max}, \quad \dots \quad (19)$$

we have finally

$$\phi_{2\max} = \frac{T}{2\pi} \dot{\phi}_{2\max} = \frac{T}{2\pi} \frac{\gamma'_2 \cdot \phi_{1\max}}{B} = \frac{\phi_{1\max}}{B} \cdot \frac{B^2}{\Theta_2} \sqrt{\frac{\Theta_1 \Theta_2}{B^2}} = \phi_{1\max} \cdot \sqrt{\frac{\Theta_1}{\Theta_2}}. \quad \dots \quad (20)$$

5. The Gyroscopic Pendulum: Two Modes of Precession.

From the system just described there is only a small step to the gyroscopic pendulum. We need merely arrange that the centre of gravity of the flywheel falls below the centre of the gimbals shown in fig. 2. The action of gravity is in this case equivalent to an elastic constraint tending to return axes 1 and 2 to a zero position. There is no difficulty involved in tackling at once the more general case, in which the constants of these elastic constraints, say of stiffness γ_1 and γ_2 , differ from each other.

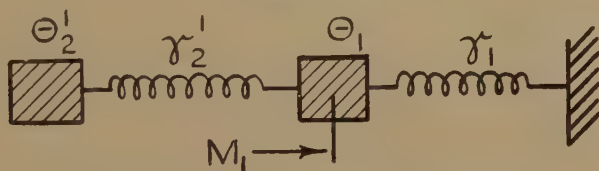
A moment M_1 applied to axis 1 "sees," then, moment of inertia Θ_1 and stiffness γ_1 combined co-resistively together with the transforms

γ_2' and Θ_2' of Θ_2 and γ_2 . As Θ_2 and γ_2 on axis 2 are combined co-resistively, these transforms themselves are combined co-yieldingly. We arrive, therefore, at the equivalent mechanical system shown in fig. 7.

There are now two frequencies of free oscillations, one in which Θ_1 moves in the same phase with Θ_2' , the other one in which it moves in anti-phase. In each case the maximum displacement of Θ_1 coincides with the maximum positive or negative displacement of Θ_2' , *i. e.* with the maximum positive or negative force exercised by γ_2' , *i. e.* with the maximum positive or negative velocity $\dot{\phi}_2$. Just as in the case discussed above, the axis of the gyrostatis describes, therefore, for each of these normal modes, a cone of elliptical or circular cross-section, and the fact that the maximum velocities just mentioned are of opposite signs in these two modes means that these cones are described in opposite senses.

The anti-phase type of movement is the one with the shorter period, the two elasticities helping each other in restoring Θ_1 to its zero position. It corresponds to the case where the precessional cone is described in such a direction that the moment M_1' created by $\dot{\phi}_2$ assists the restoring force of γ_1 . That is the case—as one can easily verify by plotting the vector of angular momentum B and its changes—if the sense in which this cone is described is the same as the rotation of the flywheel ("tendency" of the impulse vector to place itself parallel to the axis to the axis of precession).

Fig. 7.



6. The Use of the Electrical Analogy.

As already mentioned, "reciprocation" is more easily carried out in electrical diagrams than in mechanical ones. We have therefore here an additional reason to occupy ourselves with the electrical analogue of the mechanical system.

Let us translate the system illustrated in fig. 7 into an analogous electrical system according to the rules of the "natural" or "direct" analogy. In this analogy a "velocity" ω is represented by a current $i = \omega a$ and a "force" M by a voltage $e = M/a$ (a = scale factor). A mechanical impedance $\hat{z}_m = \hat{M}/\hat{\omega}$ is then represented by an electrical impedance according to *

$$\hat{z}_{el} = \frac{\hat{e}}{\hat{i}} = \frac{\hat{M}}{a} \cdot \frac{1}{\hat{\omega} \cdot a} = \hat{z}_m / a^2, \quad \dots \dots \dots (21)$$

* See the author's paper quoted above. From equation (21) it follows easily that a "mass" Θ is represented by an inductance $L = \Theta/a^2$ and a spring of constant γ by a capacity $C = a^2/\gamma$.

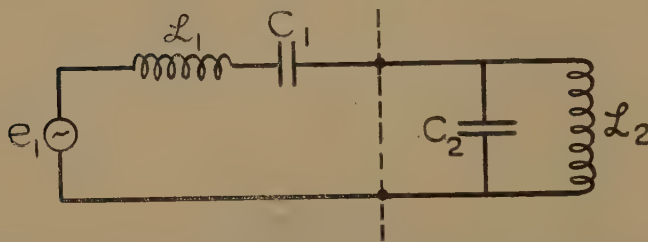
The resulting electrical network is shown in fig. 8. The dotted line in this figure separates the network in two parts, the left-hand one containing the elements connected to axis 1 of the coupler, the right-hand one those connected to axis 2. The legend to the figure gives the values of the various electrical components as they are obtained by translation from fig. 7. All those components originally connected to axis 2 of the gyrostatic coupler have, then, undergone a double process of translation, the first one according to

$$\hat{z}'_{m2} = \frac{B^2}{\hat{z}_{m2}}, \quad \dots \dots \dots (22)$$

the second one according to

$$\hat{z}_{e1} = \frac{\hat{z}'_{m2}}{a^2} \dots \dots \dots (23)$$

Fig. 8.



$$e_1 = \frac{M_1}{a}, \quad L_1 = \frac{\Theta_1}{a^2}, \quad C_1 = \frac{a^2}{\gamma_1}, \quad C_2 = \frac{a^2}{\gamma_2} = \frac{a^2}{B^2} \Theta_2 = \frac{\Theta_2}{b^2}, \quad L_2 = \frac{\Theta_2'}{a^2} = \frac{b^2}{\gamma_2}.$$

Combining (22) and (23), we get

$$\hat{z}_{e1} = \frac{B^2}{a^2} \cdot \frac{1}{\hat{z}_{m2}} = \frac{b^2}{\hat{z}_{m2}}, \quad \text{with } b^2 = \frac{B^2}{a^2} \dots \dots \dots (24)$$

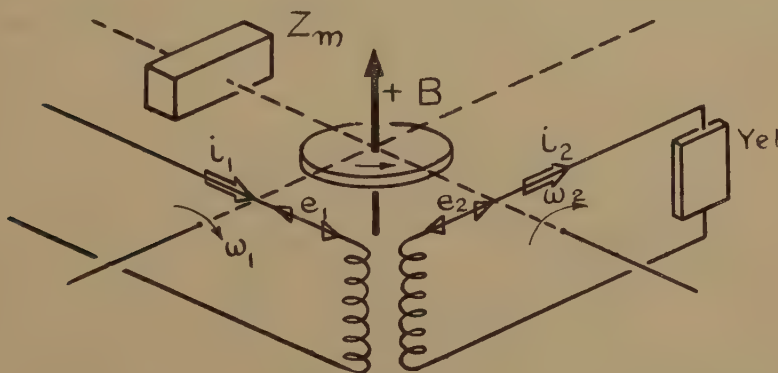
Now this is nothing else but the formula of translation according to the rules of the "inverse" analogy, in which we represent a "velocity" ω by a voltage $e = \omega b$ and a "force" M by a current $i = M/b$. As we have seen above, in the discussion of the behaviour of the gyrostatic coupler, that the translation formula for impedances contains implicitly also the rules for the translation of combinations of impedances, equation (24) shows us that we can obtain the electrical analogue of the secondary part of our mechanical system directly if we translate it according to the rules of the inverse analogy. If we had represented the transformed mechanical system of fig. 7 in terms of the inverse analogy, we would have found that the secondary part of the original mechanical system had been represented in terms of the direct analogy. Quite generally, the problem of finding the analogue of a system containing one gyrostatic coupler is then simply solved by translating either part of it according to the rules of the direct analogy, and the other part of it according to

the rules of the inverse analogy. Which analogy to use for which side is a matter of convenience.

We note that the ideal gyrostatic coupler itself is not represented at all in the resulting electrical network; it has left only a rudimentary trace of its existence in the dotted line shown in fig. 8.

We could give a stronger indication as to the existence of the gyroscopic coupler if we connected the two parts of the network by means of an ideal transformer. This has several advantages, illustrated in fig. 9: it allows us to define the direction of positive current in both parts of the network so as to coincide with the positive sense of rotation of the axis of the coupler, this sense being chosen at will. The correct sign of the relation between the two parts of the network, depending on the sense of rotation of the flywheel, is, then—quite appropriately—expressed by choosing the sign of the transformer-ratio positive or negative.

Fig. 9.



$$\begin{aligned} & +u:1. \\ i_1 &= \omega_1 a, & i_2 &= u \cdot i_1, & i_2 &= M_2/b; \\ e_1 &= M_1/a, & e_2 &= e_1/u, & e_2 &= \omega_2 b. \\ M_2 &= i_2 b = i_1 u b = \omega_1 a u b = \omega_1 B, \text{ i. e. } a u b = B. \\ y_{el} &= \frac{i_2}{e_o} = \frac{M_2}{\omega_o b^2} = \frac{z_m}{b^2}. \end{aligned}$$

Furthermore, it allows us to choose the scale factor for the representation of the secondary side at our convenience, and independent of the magnitude of the angular momentum B of the coupler (thus, we do not need a new model-circuit every time we change the value of B). For this it is necessary—see the simple derivation in the legend to fig. 9—that the transformer has the ratio $u : 1$, where a , u and b obey the relation

$$a, u, b = B. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

Choosing $a=b=1$, we get the numerically convenient relations:

$$\left. \begin{array}{lll} \text{for the primary side} & L=\Theta, & C=\Delta, \quad \hat{z}_{el}=\hat{z}_m, \\ \text{for the secondary side} & L=\Delta, & C=\Theta, \quad \hat{y}_{el}=\hat{z}_m. \end{array} \right\} \quad (26)$$

7. The Gyroscopic Ship's Stabilizer in Terms of the Electrical Analogy.

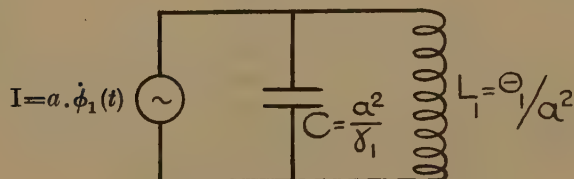
As a simple application of the above we shall apply it to a discussion of the gyroscopic ship's stabilizer. Such a stabilizer consists of a gyroscope, which is mounted, like a pendulum, in one gymbal only, the axis of which is placed athwart the ship.

We have here two oscillatory systems coupled by means of a gyrostatic coupler. The first system consists of the ship, which—on smooth water—is always restored to its upright position by its own buoyancy. This restoring couple is in first approximation proportional to the inclination of the ship, *i. e.* it is equivalent to an elastic constraint the stiffness of which can be expressed in terms of the ship's weight and metacentric height as

$$\gamma_1 = G_s \cdot h_s. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

Therefore, on smooth water the system ship + buoyant forces is equivalent to the system mass + spring shown in fig. 6. If the water surface should become inclined by an angle $\phi_1 = \phi_1(t)$ varying with time, we could imitate this action in our equivalent system by displacing the anchored point of the spring according to the same time law (this is,

Fig. 10.



perhaps, best seen if we adjust the inclination of the water surface in sudden jumps). The electrical analogue to such a system is shown in fig. 10.

The second oscillatory system consists of the stabilizing pendulum, which, with the flywheel not rotating (or with the flywheel rotating, but with the ship held in a fixed position), is just an ordinary damped pendulum. Its components ("mass" of the pendulum Θ_2 , "elastic constraint" γ_2 and viscous resistance K_2) are acting co-resistively when we try to set it into oscillation by means of a moment applied to its supporting axis.

Due to the gyrostatic action of the flywheel, this system is coupled to the oscillations of the ship about its longitudinal axis. When we try to rock the ship (Θ_1) we have also to overcome the impedance offered by the "transform" of this second system. This means that any current flowing through L_1 has also to pass through a network representing this transform, and we arrive, therefore, finally, at the electrical network shown in fig. 11.

The discussion of this network is an elementary problem of electrical engineering, and will not be further gone into. We shall merely give an

example for the retranslation of results obtained from the electrical system to the original mechanical system.

When dimensioning such a stabilizer, the angle of oscillation of the stabilizer is of importance. This amplitude is given by $\phi_2 = \int \dot{\phi}_2 dt$. As $\dot{\phi}_2 b$ appears in the electrical system as voltage e_2 across L_2 , we have $\phi_2 = \frac{1}{b} \int e_2 dt = \frac{1}{b} \cdot \Phi$, where Φ is the flux linked with L_2 . This result is, of course, only a reiteration of the general correspondence between flux and geometrical displacement according to the rules of the inverse analogy.

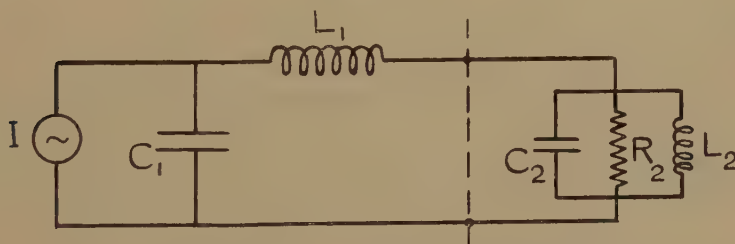
8. Further Applications :

(a) *Pseudo-regular Precession* ; (b) *Coupling due to Coriolis Forces*.

(a) *Pseudo-regular Precession*.

Let us consider a gyrostator under the influence of gravity in its state of regular precession. In this case the gyrostator axis describes at uniform

Fig. 11.



$$\begin{aligned}
 I &= a \cdot \dot{\phi}_1(t), & C_1 &= \frac{a^2}{\gamma_1}, & L_1 &= \frac{\Theta_1}{a^2}, \\
 C_2 &= \frac{\Theta_2}{b^2}, & L_2 &= \frac{b^2}{\gamma_2}, & R_2 &= \frac{b^2}{K_2}.
 \end{aligned}$$

speed a cone of circular cross-section around the vertical, the period of this orbit being determined by the condition that the externally applied moment equals the rate of change of angular momentum of the gyrostator. We shall investigate the small oscillations (nutations) of the axis of the gyrostator around this state of dynamic equilibrium. For this we place a plane through the vertical axis of this cone and the axis of the gyrostator. This plane rotates at uniform speed, and for an observer situated on it the axis of the gyrostator appears at rest (at "zero position") as long as the regular precession is maintained.

If this observer tries to alter the angle of inclination ϕ_1 of the gyrostator axis to the vertical, and has at the same time arranged suitable constraints which prevent this axis from leaving his vertical plane, then he will find, that these vertical movements are impeded only by the inertia (Θ_1) of the gyrostator (around an axis through its support and normal to

the vertical plane), and—possibly—by some elastic constraint γ_1 . In the case usually discussed in the literature this constraint is zero, the gyrostat being in indifferent equilibrium for all values of ϕ_1 .

Conversely, if the observer moves the axis of the gyrostat from its zero position normal to his vertical plane, compelling it at the same time to maintain constant inclination ϕ_1 to the vertical, he will find that these movements are again impeded only by the inertia (Θ_2) of the gyrostat (around an axis through the support of the gyrostat and lying in the vertical plane).

In the actual movements carried out by the axis of the gyrostat these constraints are absent and the two degrees of freedom just defined are then coupled by the specific action of the gyrostat. We have, then, in the case $\gamma_1=0$, a system of exactly the same kind as discussed under section 4. The free oscillations which the axis of the gyrostat can carry out around its "zero-position" are cones of elliptical or circular cross-section. If the gyrostat carries no gimbals with it, and the fly-wheel constitutes a "symmetrical top," then $\Theta_1=\Theta_2$ when, in addition to these conditions, the axis of the gyrostat passes through the point of support. This is the case usually discussed, and in this case the cone of oscillation is of circular cross-section and the period for one oscillation is given by equation (18).

The case of unequal moments of inertia, or the case where $\gamma_1 \neq 0$, leads to cones of elliptical cross-section, the discussion of which offers no special difficulties. The equivalent system is again of the type shown in fig. 6, with the elasticity being formed by the co-resistive combination of the real elasticity γ_1 and the "transformed" elasticity $\gamma'_2=B^2/\Theta_2$.

(b) *Coupling due to Coriolis Forces.*

The "concealed" and inaccessible motion which is the cause of the gyrostatic coupling terms need not always take the form of rotation of a gyrostat. As an example for dealing with such a case we consider couplings which arise from the rotation of the adopted frame of reference (couplings due to Coriolis Forces).

Let us assume that we have a mass-point (mass m) on a "roundabout" (angular velocity ω_0) which is compelled to remain inside a massless and smooth rectilinear guide. If this guide is fixed to the roundabout and we impart to the mass-point an oscillatory velocity \hat{v}_1 (angular frequency Ω) this guide experiences a side-thrust

$$\dot{P}_{az}=2\omega_0 m \hat{v}_1 = B \hat{v}_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

with

$$B=2\omega_0 m \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

arising from the Coriolis acceleration associated with this velocity.

The force \hat{P}_1 required to establish this velocity is

$$P_1=j\Omega m \hat{v}_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

i. e. the "input impedance" of the mass-point has its normal value,

Now let the guide be so arranged that it can move sideways. In these circumstances it can act as a "source of force" which can be used to drive some mechanism. Suppose the sideways velocity of the guide is \hat{v}_2 (also periodically varying), then we have to increase \hat{P}_1 by an amount

$$\hat{P}_{c1} = B\hat{v}_2. \quad (31)$$

If
$$v_2 = \frac{P_{c2}}{\hat{z}_2} \quad (32)$$

(where \hat{z}_2 has to include the "internal" impedance $j\Omega m$ of this "source of force"), then

$$\hat{P}_u = \hat{P}_1 + P_{c1} = j\Omega m + \frac{B^2}{\hat{z}_2} \quad (33)$$

If ω_0 is very large we may have a finite value of B with $m \rightarrow 0$, and then the first term in (33) can be neglected. This case corresponds exactly to that of the ideal gyrostatic coupler discussed in section 2. The usual case, with m finite, corresponds to the case of the real coupler discussed under section 4. If there is no further loading applied to the coupler, then $\hat{z}_2 = j\Omega m$, i. e. P_{1u} "sees" a transform $c' = B^2/m$ and we have a period of free oscillation

$$T = 2\pi \sqrt{\frac{m}{c'}} = 2\pi \sqrt{\frac{m \cdot m}{B^2}} = \frac{\pi}{\omega_0} = \frac{T_0}{2}, \quad (34)$$

where T_0 is the time required for one rotation of the roundabout.

This "free oscillation" consists of a circular path, as can be established by the same method as used in section 4, and which is, of course, nothing else but the path at which the centrifugal force just balances the Coriolis force.

An external force \hat{P}_1 of just this periodicity then "sees" zero input impedance.

The case just discussed, especially the last result, is the mechanical analogue of an electrical case of considerable technical importance—the movements of an electron in a magnetic field—but this cannot be gone into further on this occasion.

9. Systems with more than one Gyroscopic Coupler.

The use of the methods here described is not limited to systems containing only one gyroscopic coupler. As an example we shall apply it to a system containing two gyroscopes which are elastically coupled to each other and to "ground." Such a system has been investigated by traditional methods by K. Magnus*.

In this arrangement the gymbal axes 1 and 2 of the first gyroscope are parallel to the axes 3 and 4 of the second gyroscope. Each gymbal ring is elastically constrained to "ground" (elastic constants $\gamma_1, \gamma_2, \gamma_3, \gamma_4$), and, in addition, there are elastic couplings between gymbal rings 1 and 3 and 2 and 4 (coupling moments $M_{c1} = \gamma_5(\phi_3 - \phi_1)$ and $M_{c2} = \gamma_6(\phi_4 - \phi_2)$).

* *Ingenieur Archiv*, vol. ix, p. 178 (1938).

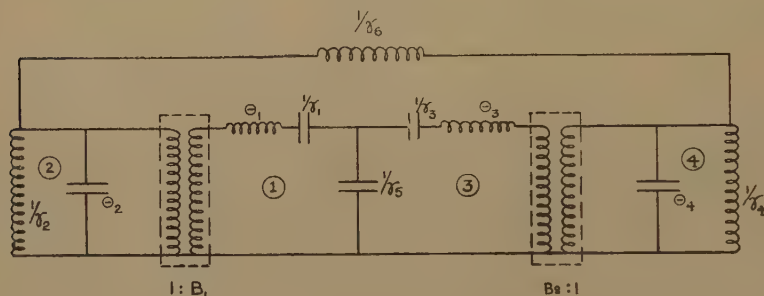
To arrive at the electrical analogue of this system we represent the elements connected to axes 1 and 3 according to the natural analogy, and the elements connected to axes 2 and 4 according to the inverse analogy. The resulting network (with $a=1$, $b=1$) is shown in fig. 12.

10. *The Structure of Systems to which the Present Method is applicable :
Remarks concerning the Behaviour of such Systems.*

From the examples given it is seen that the present method (transformation by reciprocation) is applicable to all those cases where we can divide the whole system uniquely into subsidiary systems of two classes (which we may appropriately call the “ a ” and the “ b ” classes—having in mind their subsequent representation as electrical networks according to the direct or the inverse analogy), so that whenever we cross a gyrostatic coupler we cross from one class into the other class.

In carrying out this “marking” process it is immaterial if we start with an “ a ” or “ b ” designation. The marking process can be carried

Fig. 12.



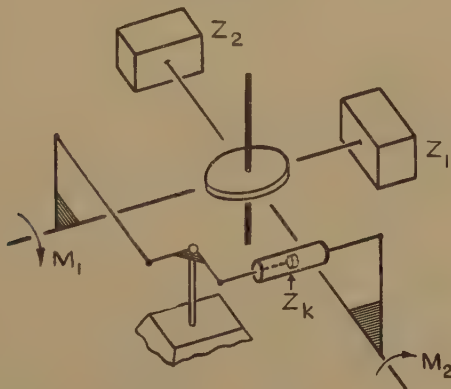
out, of course, either on the pictorial representation of the system or on the set of differential equations describing its behaviour. We need only observe which of the two different types of coupling (“ordinary”=symmetrical and “gyroscopic”=asymmetrical) leads us from one co-ordinate to the “next” one. The two alternatives of starting result in two different networks, the one being the dual counterpart of the other. However, both are simple, passive, electrical networks.

Now from this last remark we can immediately draw a few conclusions concerning the behaviour of such systems. Suppose we select two pairs of points of the original mechanical system, both pairs belonging to subsystems of the same class. Then an observer who has access only to these four “terminals” will not be able to detect the presence of gyrostatic couplers in the system. His observations are only sufficient to determine the passive elements of a 4-terminal system capable of simulating the actual system, no matter how complicated the actual system may be. This follows immediately from the fact that we can contract every passive electrical network between four input terminals into an equivalent 4-terminal network which in the usual case (input applied only to either

the one or the other pair of terminals and not between members of different pairs) contains only three passive elements. We need only to choose "b" markings for those subsidiary systems in which our input and output "points" are situated in order to obtain electrical networks in which these points are represented as "junctions" in order to make the analogy very obvious.

If our mechanical pairs of terminals fall into different classes, the contraction of the electrical network into an equivalent 4-terminal network can still be carried out; so can the contraction of the corresponding transformed mechanical network. However, the input to one side of the transformed mechanical network is the "inverse" of the actual input, and, if we wish to obtain a system to which we can apply the actual input we have to add to the resulting 4-terminal network an ideal gyrostatic coupler. Thus we can replace the actual system, however complicated it is, by a passive 4-terminal network followed by an ideal gyrostatic coupler. This last equivalent system can be modified in a

Fig. 13.



number of ways, as we can always transfer one or more of the components of the 4-terminal network to the other side of the ideal gyrostatic coupler (under reciprocity, of course), in the same way as we transfer components to the other side of an ideal transformer.

11. Systems with Asymmetrical Impedance Properties.

If we carry out the marking process described in the last section we may encounter one or more components which, when approached by one route may get an “*a*” marking, and when approached by another route would get a “*b*” marking. In this case our method breaks down, as we can neither say how we should represent these components in the electrical analogue nor how they should appear in the transformed mechanical system.

Systems of this kind, however, possess the remarkable property that certain of their transfer impedances are neither symmetrical nor skew symmetrical. The simplest example for such a system is given in fig. 13,

which shows schematically an ideal gyrostatic coupler to which has been added between the axes 1 and 2 a further coupling link of impedance z_k . The governing equations for such a system are

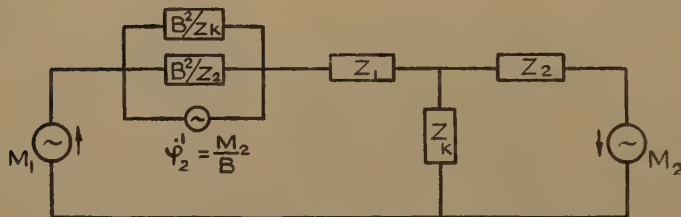
$$\left. \begin{aligned} \omega_1(z_k + z_1) + \omega_2(B - z_k) &= M_1, \\ -\omega_1(B + z_k) + \omega_2(z_k + z_2) &= M_2. \end{aligned} \right\} \quad (35)$$

It is obvious that an electrical analogue to such a system could be constructed by the use of valves (though no amplifying action could be obtained in this case, as the mechanical system does not include a source of energy). Of interest is, however, that it is possible to derive a representative electrical circuit which contains—apart from the elements representing M_1 and M_2 —only passive elements.

Eliminating ω_2 by the aid of the second of the equations (35) we get

$$\left. \begin{aligned} M_1 &= \omega_1(z_k + z_1) + \omega_1 \frac{(B + z_k)(B - z_k)}{z_k + z_2} + \frac{M_2(B - z_k)}{z_k + z_2}, \\ &= \omega_1 \left\{ z_1 + \frac{z_k z_2}{z_k + z_2} + \frac{B^2}{z_k + z_2} \right\} + \frac{M_2}{B} \cdot \frac{B^2}{z_k + z_2} - M_2 \frac{z_k}{z_k + z_2}. \end{aligned} \right\} \quad (36)$$

Fig. 14.



This equation is interpreted by the circuit shown in fig. 14. We notice that the voltage source representing the driving moment M_1 “sees” the elements of the secondary side at the same time in two different ways: once as if there existed only a symmetrical coupling impedance z_k , and once as if there existed only the skew symmetrical coupling impedance B .

In this analogous electrical system ω_1 is represented by the current delivered from M_1 ; the secondary driving moment is represented once by the voltage source M_2 , and once by the current source $\phi_2 = \frac{M_2}{B}$. ω_2 has to be obtained by several operations as follows: we can rewrite the second of the equations (35)

$$\omega_2 = \omega_1 \cdot \frac{B^2}{z_k + z_2} \cdot \frac{1}{B} + \frac{M_2 + \omega_1 z_k}{z_k + z_2}, \quad (37)$$

and from this form we see that ω_2 is represented in part by the voltage drop caused by the flow of ω_1 through the inverted system $B^2/(z_k + z_2)$ (this voltage drop converted into current by multiplication with $1/B$), and in part by the current which circulates in the secondary mesh due to the action of the voltage source M_2 and of the primary current ω_1 .

XXXIX. *A Centrifugal Method of Measuring the Surface Tensions and Interfacial Tensions of Liquids in Capillary Tubes.*

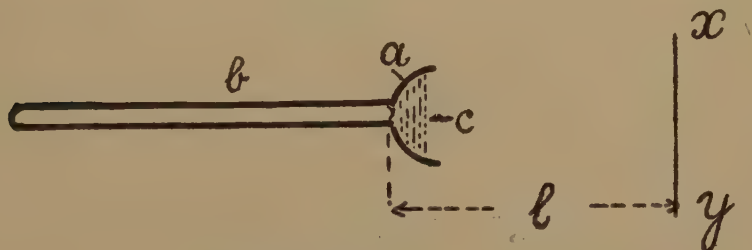
By W. MEYERSTEIN, M.D.* (Department of Physiology in the University of Birmingham) and J. D. MORGAN, D.Sc.†

[Received March 28, 1944.]

Introduction.

PRIOR to the investigation forming the subject of this paper, one of the writers (W. M.) had devised a simple means and method of filling hæmatocrit capillary tubes‡. The means consists of a small cup *a* (fig. 1) formed on or secured to the open end of a capillary tube *b* (the other end being closed), the cup having sufficient capacity to hold a small quantity of the liquid *c* to be transferred to the tube.

Fig. 1.



The method consists in spinning the tube in a centrifuge about an axis as *x, y*.

At a particular speed the liquid in the cup passes into the tube suddenly. This speed depends on the nature of the liquid, the diameter of the bore of the tube, and the distance *l* from the axis of rotation. Some preliminary experiments showed that for a given liquid and a given distance *l*, the speed of rotation varied inversely as the diameter of the bore. It was also noticed that the tube could be emptied by reversing the position of the open end relatively to the axis of rotation. As in filling, emptying occurred suddenly at a particular speed.

From his experience with his hæmatocrit tubes the writer (W. M.) conceived the idea that his method might provide the basis of a new method of measuring the surface tensions of liquids, and particularly

* Caroline Harrold Research Fellow.

† Communicated by the Authors.

‡ Proceedings of the Physiological Society, 10th May, 1942, in 'Journal of Physiology,' vol. 101.

interfacial tensions. The exploration of this idea was undertaken jointly by both writers, and the results form the subject of the present paper.

Explanation of Action.

The action on which the method depends can readily be understood from the following :—

Suppose a tube *a* (fig. 2), closed at one end and open at the other and mounted radially on a horizontal rotary table (axis of rotation *x, y*), has a quantity of liquid located near the open end of the tube. The inner end of the liquid is bounded by the meniscus *b*. If the meniscus

Fig. 2.

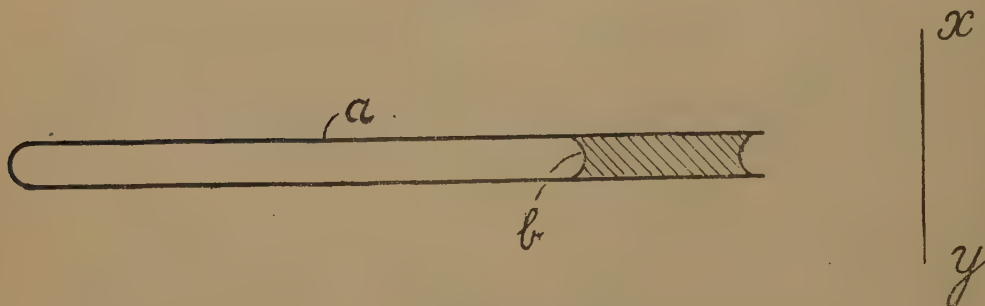
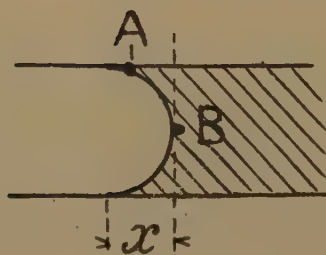


Fig. 3.



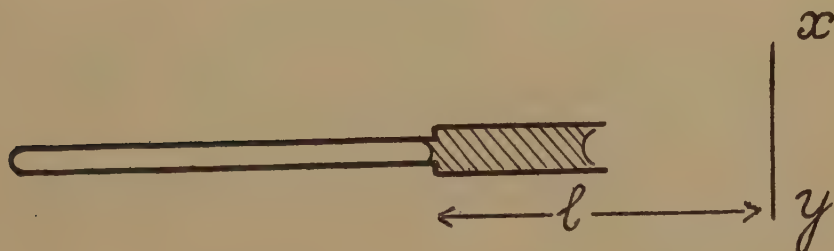
were flat, rotation (at a speed appropriate to the action under consideration) would not cause the liquid to flow to the other end of the tube. The liquid would merely act like a plug or piston, and the effect of rotation would only be a displacement of liquid towards the closed end, causing compression of the air contained in the tube. On stopping rotation the liquid would return to its initial position under the pressure of the compressed air. The fact that (when the meniscus is curved) flow of liquid occurs along the tube from the open to the closed end at an appropriate speed of rotation, indicates that the effect is attributable solely to a condition associated with the meniscus, and the following is a reasonable supposition.

If the pressure at *A* (fig. 3) due to centrifugal force exceeds that at *B* by an amount equal to the pressure ($4T/d$) due to surface tension, the

latter pressure at A will be neutralized, and a slight increase of speed will cause the liquid to flow as a film along the inner surface of the tube. Experiments have proved this supposition to be correct.

In the case of a tube of uniform diameter, as shown in fig. 2, transfer of liquid does not occur suddenly, as in the Meyerstein hæmatocrit tube, but gradually. For a given speed there is a position of equilibrium of the meniscus. Increase of speed is followed by the transference of more liquid until the meniscus takes up a new position corresponding to the increased speed. Only when the speed is made sufficiently high is the whole of the liquid transferred. But when a tube consisting of two parts of different diameters is used, as indicated by fig. 4, and the liquid is placed in the larger part, the transference to the capillary portion occurs suddenly and completely at a particular speed, the event being often accompanied by an easily audible gurgle. When the position of the tube is reversed, emptying occurs suddenly and completely at the appropriate speed. It will, of course, be understood that for both filling

Fig. 4.



and emptying, the meniscus is initially situated at or only a short distance within the open end of the bore of the capillary portion of the tube.

Theory.

On the assumption that the meniscus is hemispherical, the above explanation leads to the following formula, which is applicable to both filling and emptying, provided that the action is of the above-mentioned kind which occurs suddenly at a particular speed.

$$T = \frac{1}{2} \rho d^2 \pi^2 n^2 l, \quad \dots \dots \dots (1)$$

where T = surface tension.

ρ = density of liquid.

d = diameter of bore of capillary tube in cms.

n = rate of rotation in revs./sec.

l = distance of meniscus from axis of rotation in cms.

Where the transfer of liquid in the tube occurs gradually (as in the case illustrated by fig. 1), the formula (2), to be mentioned later, applies

Apparatus.

In our experiments both of us used essentially similar apparatus. This consisted of a rotary disk mounted horizontally on the upper end of a vertical spindle driven by an electric motor, the speed of the latter being controlled by a variable resistance. The capillary tube was secured in a radial position on the disk over a narrow radial slot in the disk, and an electric lamp was mounted beneath the disk so that it could illuminate the tube. Over the disk was mounted a stationary screen having a narrow slot, through which movement of the liquid in the tube could be observed. For measuring the speed of the disk W. M. used a counter and stop-watch and J. D. M. used a Haseler Indicator, this latter being a combined direct reading counter and watch. Accurate measurement of speed presented the major difficulty encountered by us. It is a disadvantage of the method that (other factors being constant) T depends on n^2 , as a small error in the measurement of n may result in a large error in the evaluation of T . Also it is necessary to avoid heating of the tube either by the lamp or the motor.

Test of Theory: Filling Method.

To test the theory, liquids of known ρ and T were used. The results of calculations and observations with a few liquids are given in the following table, where speeds are recorded as revs./min. :—

Liquid.	T.	ρ .	$d(\text{cm.})$	$l(\text{cm.})$	Cal. r.p.m.	Obs. r.p.m.
Water	73	1	0.15	7.5	556	558
Paraffin oil	26	0.8	372	370
Benzene	29	0.9	372	370

This method does not appear to be applicable to all liquids, as the very mobile ones run into the capillary tube of their own accord. It may be that this difficulty could be met by using capillary tubes of sufficiently small bore. It will be noticed, however, that in the examples quoted there is sufficiently good agreement to warrant the conclusion that the method is attractive and the theory correct. In calculating the above results (and those given later) only approximate values of T and ρ have been used, and the observed speeds and measurement of the diameter of the capillary tube bore may contain small errors. But our experience supports the belief that with improved apparatus a higher degree of accuracy could be attained.

Emptying Method.

The theory requires that emptying shall occur at the same speed as filling (l and d being the same in both cases). In the emptying tests capillary tubes as shown in fig. 5 were used.

With water as the test liquid, it was noticed that only occasionally did the discharge occur at or near the calculated speed. More often the discharge occurred at a higher speed, corresponding nearly to $\sqrt{2n^2}$ (where n is the calculated filling speed). The investigation of this anomaly led to an interesting and possibly valuable result. It was noticed that, prior to emptying, an air-bubble was often formed at the open end of the tube, as shown by *a*, fig. 6, and that this bubble travelled more or less rapidly to the closed end.

Further, it was noticed that the contents were discharged at the calculated value of n only when no bubble was formed, and that when a bubble occurred the meniscus at the open end of the tube was flattened from its initial hemispherical form sufficiently to account for the observed higher

Fig. 5.

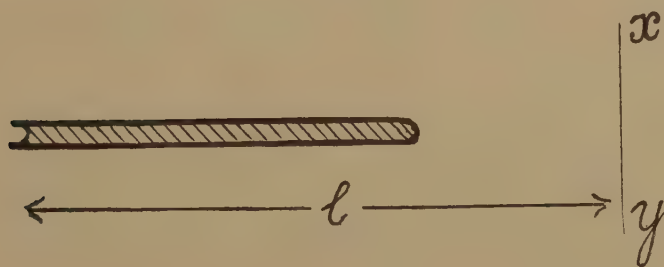
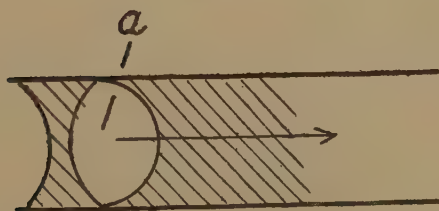


Fig. 6.

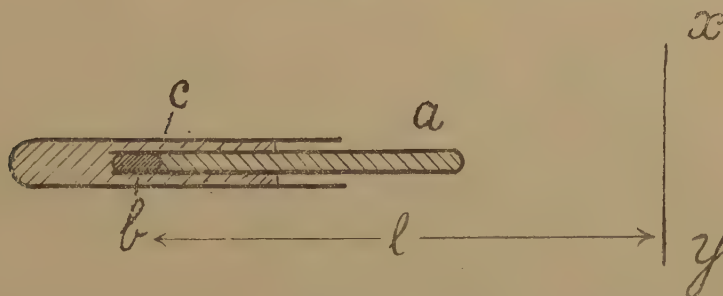


speed required to effect the discharge. Moreover, when other liquids of lower T were used (chloroform, benzene, paraffin oil, etc.), it was noticed that the bubble was nearly always formed, and that discharge at the proper speed without the previous occurrence of the bubble was a relatively rare event. A bubble having been formed, the emptying speed was about $\sqrt{2n^2}$, where n is the filling speed. Also it was observed that in all cases the formation of the bubble took place at a speed considerably lower than the calculated filling or emptying speed n .

The bubble phenomenon seemed to us to have an important bearing on our investigations, and it was therefore given special attention. It was immediately obvious that the speed at which the bubble appeared did not comply with formula (1), and experiments with capillary tubes of different bore-diameters, and with different liquids, disclosed that its

sprung on to the inner tube and arranged to abut against the open end of the outer tube. On spinning the tubes at the appropriate speed the liquid *b* will be seen to move slowly across the field of view. In this case formula (1) applies, subject to the condition that ρ now represents the difference between the densities of the two liquids. As in the bubble experiments, the results obtained were very consistent, and the rate of movement of the liquid *b* afforded a valuable aid in the accurate determination of *n*. This method is analogous to the bubble method and may

Fig. 7.



be conveniently described as the drop method. The results of some experiments with the drop method are given in the following table :—

Liquids.	ρ diff.	Interface. T.	<i>d</i> (cm.)	<i>l</i> (cm.)	Cal. r.p.m.	Obs. r.p.m.
Water-chloroform .	0.5	33	0.2	10	342	350
„ -benzene ...	0.1	34	780	790
„ -paraffin oil .	0.2	48	660	650

Comments.

As the drop method of measuring interfacial tension appears to be analogous to the bubble method for single liquids, it may appear strange that the same formula does not apply to both. It must be remembered, however, that formula (1) depends on the assumption that the meniscus is hemispherical, and that, in consequence, the length of liquid *x* between the point A (fig. 3) and a plane (at right angles to the tube) containing the point B is $d/2$. In the interface experiments this condition appeared to be satisfied, but in the bubble experiments the meniscus (so long as it occupied a distance of about *d* or rather more from the open end of the tube) first became elongated and subsequently its lip turned inward to form the bubble. The elongated meniscus was probably of paraboloidal form, and its length *x* corresponded to the expression $x=cd^2$. The quantity d^2 accounts for the quantity d^3 in (2), and the evaluation of *c* showed that for all the tubes and liquids used it was as near 10 as could be measured.

In carrying out the interfacial tests by the drop method it was sometimes observed that the transference of the liquid *b* occurred at a speed very much higher than the proper speed, the observed speed being often as high as $\sqrt{2}n^2$ (n being the proper speed), as in the emptying experiments above described. Here again it was noticed that in this condition a tiny droplet of liquid had been broken off the inner end of the main drop *b*, leaving a flattened meniscus. Moreover, when the main portion of the liquid *b* moved across the field of view, it did so very quickly, and this was a sure sign that an undesired condition had entered the experiment.

Conclusion.

The above-described methods, and particularly the bubble and drop methods, appear to be capable of giving accurate and consistent results, and it is submitted that they provide a useful contribution to the art of measuring surface tensions and interfacial tensions, especially where only small quantities of liquid are available, as in some physiological investigations. Whilst we have described more particularly the bubble and drop methods as being probably the most satisfactory for the liquids used by us, other variants are practicable, and may prove to be equally good or even more suitable in special cases.

XL. *An Investigation of the Existence of "Magnetic" Currents.*

By D. M. MILLEST, B.Sc.*

[Received May 15, 1944.]

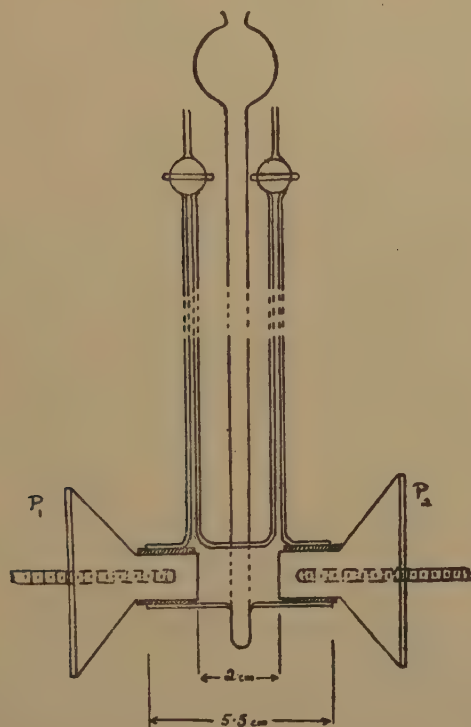
ACCORDING to a brief account published in the daily press of January 17th, Ehrenhaft claims to have established the existence of a "pure magnetic current." This conclusion is based on the observation that, when the poles of a magnet were dipped into acidulated water, oxygen and hydrogen were evolved at both poles, but "a greater quantity of oxygen was found in every case to have come from the north pole."

In Ehrenhaft's experiments, the direct action of the acid on the material of the poles would lead to the production of hydrogen, while the presence of oxygen at both poles might be due to the liberation of air dissolved in the acid. Any decomposition which might take place as a result of the action of the magnetic field must therefore be superimposed on these two effects, and would be very difficult to identify unambiguously. In view of the difficulty of accounting for the "magnetolysis" of acidulated water, and of the important influence which such an effect, could it be substantiated, would have on theories of the nature of electricity and magnetism, it seems to be desirable to

* Communicated by Frank Horton, F.R.S.

investigate the effects produced in circumstances such that a direct action of the acid on the material of the pole-pieces is prevented, and the liberation of dissolved air is reduced to a minimum.

The apparatus designed in order to meet these conditions is shown in the diagram. The iron pole-pieces, P_1 and P_2 , were cleaned and heavily coated with tin, where they would otherwise have been in contact with the acid. They were then screwed onto the iron core of the electro-magnet, and a current of 10 ampères was passed through the coils, producing a field of 3,000 gauss at the centre of the region between the poles. The liquid in the voltameter was ordinary battery acid, diluted



with an equal volume of distilled water. It was boiled before use, to drive off the dissolved air.

The procedure usually adopted was to pass the current through the magnet coils for about 8 hours, and then to leave the apparatus overnight with the current zero. On the following day, the direction of the current was reversed, the pole which had been north, now becoming a south pole.

After the first day, during which gas was evolved at neither pole, whether the magnetizing current was on or off, the acid began to attack the metal of the right-hand pole. Observations were therefore continued on the phenomena taking place at the left-hand pole-piece. No gas was ever detected here in the absence of a magnetizing current. The

results obtained when the field of 3,000 gauss was established are summarized in the following table:—

Left-hand pole south.		Left-hand pole north.	
Time of passage of current.	Volume of gas collected.	Time of passage of current.	Volume of gas collected.
Hours.	c.mm.	Hours.	c.mm.
11.5	Zero.	—	—
—	—	7.0	3
—	—	8.5	Zero.
9.0	Zero.	—	—
—	—	8.5	<1
9.0	Zero.	—	—
—	—	9.5	<1
8.0	3	—	—
—	—	8.0	Zero.

These results show that gas was evolved on only one occasion out of the four periods when this pole was a south pole. When the polarity was reversed, gas did not appear on two occasions. In the remaining three periods, small quantities of gas were detected.

It is clear at once that the liberation of gas does not take place in the regular and consistent way which would be expected, if its production was due to a genuine effect of the magnetic field. The erratic nature of the phenomenon is emphasized by the fact that there is no correlation between the volume of gas evolved, and the time during which the magnetic field existed. During one experiment, for example, 3 c.mm. were liberated in 7 hours, while in another, less than 1 c.mm. appeared after 9.5 hours, the intensity of the magnetic field being the same in the two cases.

These results indicate that the accumulation of gas is due to the liberation of dissolved air from the liquid. A thermometer, inserted into one of the pole-pieces, registered a temperature of 70° C. when the magnetizing current had been on for several hours; so the gradual liberation of some of the air dissolved during the process of filling the voltameter might be expected to take place. The detection of a small bubble of gas between the pole-pieces, in the horizontal arm of the voltameter, during some of the periods when the magnetizing current was on, suggests that this explanation is correct, and that gas is liberated from the bulk of the acid, and not exclusively at the poles.

While this investigation was being carried out, an article by Kendall* was published, describing the results of rather similar experiments. He showed that no gas was liberated when the pole-pieces of an electro-magnet were coated with a film of paraffin wax and dipped into a

* 'Nature,' cliii. p. 157, February 5th, 1944.

solution of ferrous chloride. In a second experiment, the bare pole-pieces of an electromagnet were immersed in dilute hydrochloric acid. In this case only about 0.5 per cent. of the gas liberated at either pole was oxygen, the volumes of oxygen collected at the north and south poles differing from each other by an amount less than the experimental error in analyzing the gaseous mixture.

Now both these observations cast considerable doubt on the existence of a "magnetolytic" effect, but they do not completely exclude the possibility. It is very difficult to understand why such an action should take place, and correspondingly difficult to envisage the best conditions for its manifestation, assuming that it does exist. In the absence of any indications from theoretical considerations as to the most favourable arrangements, all that can be done is to try experimental conditions, departing not too far from those in which the effect was first reported to have been observed. In the experiment with the uncoated pole-pieces, Kendall collected the gas from each pole, separately, by placing a thin strip of mica between them. This was necessary since their distance apart was only 1.5 mm. The absence of a significant difference between the oxygen content of the two samples of gas might possibly have been due to the presence of this barrier in the region where the intensity of the magnetic field was greatest. The negative results of the experiment with the wax-coated pole-pieces might perhaps imply that a continuous electric circuit, as well as the presence of a magnetic field, is necessary for the production of "magnetolysis." If it is true that, in virtue of previously unsuspected magnetic properties, the ions can be made to travel through the liquid by the presence of a magnetic field, it seems likely that, for continuous decomposition to take place, they should be able to give up their electric charges on reaching the poles. If they cannot do this, a back E.M.F. would be set up, counteracting any motion which the magnetic field might produce. Neutralization of the ions cannot take place if a layer of wax separates the poles from the liquid.

In the apparatus used in the present investigations, the ions could travel freely across the liquid, and would encounter a metal surface at the boundary. It seemed, therefore, that the conditions were more favourable than those used by Kendall for detecting any decomposition due to the magnetic field, and that, despite this additional evidence of the improbability of any such action, it was worth while to complete the experiments.

But even in these conditions, no trace of "magnetolysis" was found to take place. Kendall concluded that the small percentage of oxygen collected in one of his experiments, was due to the liberation of air dissolved in the acidulated water. The results of the present investigation are in agreement with this conclusion.

Royal Holloway College,
Englefield Green,
Surrey.

*XLI. On Helmholtz's Electrodynamical Potential and the Induction
Coefficients of Unclosed Currents in Quasi-stationary Processes.*

By A. DA SILVEIRA, Professor of Physics at Instituto Superior Tecnico,
Lisbon *.

[Received July 25, 1943.]

THE problems concerning the electro-dynamical potential and the induction coefficients of unclosed currents in quasi-stationary processes are not generally treated in books on the theory of electricity, or are treated in a not very satisfactory manner. The present treatment is an elaboration of that given by Planck, but is not subject to the severe criticisms of the one presented by this author †.

We shall begin by establishing some formulæ necessary for the calculation of the magnetic energy W_m^* of a system of currents, or of its electro-dynamical potential $-W_m^*$. It is in the deduction of these formulæ that the known treatments fail.

When a conductor is traversed by currents variable in time, it is known that there exists in the interior of the conductor a density of electricity ρ variable in time :

$$\frac{\partial \rho}{\partial t} = -\text{div } \vec{J} = \frac{1}{4\pi} \text{div } \frac{\partial \vec{D}}{\partial t}, \quad (1)$$

and that the conductor's surface is covered by a layer of electricity of density σ , also variable in time :

$$\frac{\partial \sigma}{\partial t} = -\vec{J}_n = \frac{1}{4\pi} \frac{\partial}{\partial t} (\vec{D}_n + \vec{D}_{n_i}). \quad (2)$$

In these formulæ \vec{J} designates the density of current and \vec{D} the electrical-displacement ; n_e and n_i designate the semi-normals exterior and interior to the conductor.

The charges $\partial \rho / \partial t$ and $\partial \sigma / \partial t$ produce an electrical field \vec{E} variable in time. If the phenomenon is quasi-stationary we can admit that there exists at every instant a scalar potential-function V from which derives the vector \vec{E} , and that this function is calculated as if the charges in each point of the conductor conserve the values they possess at the considered instant. This means, we suppose, that the field accompanies instantaneously, on all extension, the variations of charges. That is the hypothesis of current electrotechnics.

This put down, let us consider in a vacuum a system of conductors, closed or open, but perfect, that is free of electrical polarization and

* Communicated by the Author.

† M. Planck, 'Theory of Electricity,' p. 193.

without magnetic polarization (for simplicity's sake). The dielectric constant and the magnetic permeability are respectively ϵ_0 and μ_0 at any point of the space. Let us examine first the scalar potential function V . In any point $Q(x', y', z')$ of the space this function has, at any instant, the value

$$V_Q = \frac{1}{\epsilon_0} \int \frac{\rho}{r} \delta\tau + \frac{1}{\epsilon_0} \int \frac{\sigma}{r} \delta\Sigma, \quad (3)$$

where τ designates the volumes traversed by the currents and Σ the boundary surfaces. Therefrom we deduce

$$\dot{V} = \frac{1}{\epsilon_0} \int \frac{1}{r} \frac{\partial \rho}{\partial t} \delta\tau + \frac{1}{\epsilon_0} \int \frac{1}{r} \frac{\partial \sigma}{\partial t} \delta\Sigma, \quad (4)$$

In virtue of (1) and (2), and making use of the development of $\text{div} \left(\frac{\vec{J}}{r} \right)$, this equation takes the form

$$\dot{V} = \frac{1}{\epsilon_0} \int \left(\vec{J}, \text{grad}_A \frac{1}{r} \right) \delta\tau, \quad (5)$$

where $A(a, b, c)$ designates a generic point of the space τ traversed by the currents. This integral (5) can be supposed extended to the whole space v , because $\vec{J}=0$ outside of τ . Let us remark that \dot{V} vanishes at infinity just as $1/r^2$, which could be considered evident *a priori*, since from (1) and (2)

$$\int \frac{\partial \rho}{\partial t} \delta\tau + \int \frac{\partial \sigma}{\partial t} \delta\Sigma = 0$$

at every instant.

Now let $P(x, y, z)$ be a potentiated point of the space v and let us consider the Mathieu's second potential, putting

$$\psi_P = \frac{1}{\epsilon_0} \int r \frac{\partial \rho}{\partial t} \delta\tau + \frac{1}{\epsilon_0} \int r \frac{\partial \sigma}{\partial t} \delta\Sigma, \quad (6)$$

Differentiating with respect to x , and putting $\vec{D} = \epsilon_0 \vec{E} = -\epsilon_0 \text{grad } \dot{V}$ in (1) and (2), we obtain

$$\frac{\partial \psi}{\partial x} = -\frac{1}{4\pi} \int \frac{\partial r}{\partial x} \text{lap } \dot{V} \delta\tau - \frac{1}{4\pi} \int \frac{\partial r}{\partial x} \left(\frac{\partial \dot{V}}{\partial n_i} + \frac{\partial \dot{V}}{\partial n_e} \right) \delta\Sigma, \quad (7)$$

To give this expression a more suitable form, we shall consider first the space τ occupied by the conductors limited by Σ , and afterwards the space τ' limited internally by the surfaces Σ and externally by a spherical surface S with centre P and of infinitely great radius. Let us apply Green's identity to the functions $\partial r / \partial x$ and \dot{V} in the interior of each one of these spaces; let us add the results. Noting that \dot{V} and $\partial r / \partial x$ are continuous on Σ , and \dot{V} is of the order of $1/r^2$ on S , we obtain

$$\frac{\partial \psi}{\partial x} = -\frac{1}{4\pi} \int \dot{V} \text{lap } \frac{\partial r}{\partial x} \delta v = -\frac{1}{4\pi} \int \dot{V} \frac{\partial}{\partial x} \text{lap } r \delta v.$$

As $\text{lap } r = 2/r$, we can therefore write

$$\text{grad}_P \psi = -\frac{1}{2\pi} \int \dot{V} \text{grad}_P \frac{1}{r} \delta v. \quad (8)$$

Let us utilize the identity

$$\dot{V} \text{grad}_P \frac{1}{r} = \text{grad} \left(\frac{\dot{V}}{r} \right) - \frac{1}{r} \text{grad} \dot{V},$$

noting that \dot{V} is defined in the generic point Q of the space v . As long as the point Q does not coincide with P, $\text{grad}_P \dot{V} = 0$; when P is in the interior of δv , the contribution of this element is infinitesimal. We can therefore write

$$\text{grad}_P \psi = -\frac{1}{2\pi} \int \text{grad}_P \left(\frac{\dot{V}}{r} \right) \delta v. \quad (9)$$

Now the identity

$$\int \text{grad}_Q \left(\frac{\dot{V}}{r} \right) \delta v = \int \frac{\dot{V}}{r} \vec{n} \delta S$$

gives

$$\text{grad}_P \psi = -\frac{1}{2\pi} \int \frac{\text{grad}_Q \dot{V}}{r} \delta v, \quad (10)$$

in virtue of the behaviour of \dot{V} on Σ and S . Let us observe that in (9) the operators \int and grad are not commutable; in other words, from (9) we cannot conclude, putting apart an additive constant, that

$$\psi = -\frac{1}{2\pi} \int \frac{\dot{V}}{r} \delta v, \quad (11)$$

as it appears in certain authoritative works, because this integral, extended over all space, is not convergent. Nevertheless equation (10) is exact, and we have

$$\text{lap } \psi = 2\dot{V}. \quad (12)$$

In the second place let us examine the potential-vector \vec{A}^* . In a quasi-stationary process we have

$$\vec{A}^* = \frac{\mu_0}{c} \int \frac{1}{r} \left(\vec{J} + \frac{\vec{D}}{4\pi} \right) \delta v, \quad (13)$$

or

$$\vec{A}^* = \vec{A} + \frac{\mu_0}{4\pi c} \int \frac{\vec{D}}{r} \delta v, \quad (14)$$

designating the potential-vector of currents of conduction by \vec{A} . We must have, according to Maxwell,

$$\text{div } \vec{A}^* = 0. \quad (15)$$

The results obtained previously permit us to transform in (14) the term relative to the displacement-current. In fact, we can put

$$-\frac{\mu_0 \epsilon_0}{4\pi c} \int \frac{\text{grad}_Q \dot{V}}{r} \delta v = \frac{\mu_0 \epsilon_0}{2c} \text{rad}_P \psi,$$

so that
$$\vec{A}^* = \vec{A} + \frac{\mu_0 \epsilon_0}{2c} \text{grad}_P \psi, \quad \dots \quad (16)$$

From here, and considering (15) and (12), we can deduce

$$\text{div } \vec{A} = -\frac{\mu_0 \epsilon_0}{2c} \text{lap } \psi = -\frac{\mu_0 \epsilon_0}{c} \dot{V}. \quad \dots \quad (17)$$

The magnetic induction \vec{B} and the magnetic field \vec{H} are given by

$$\vec{B} = \mu_0 \vec{H} = \text{rot. } \vec{A}^* = \text{rot. } \vec{A}, \quad \dots \quad (18)$$

whence the important conclusion: in the case of unclosed currents, in quasi-stationary processes, the magnetic field \vec{H} and the induction \vec{B} are calculated from the potential-vector \vec{A} of the conduction currents, without it being necessary to take account of the currents of displacement, in spite of $\text{div } \vec{A} \neq 0$.

This said, let us consider the magnetic energy of the system :

$$W_m^* = \frac{1}{8\pi} \int (\vec{H}, \vec{B}) \delta v. \quad \dots \quad (19)$$

As there are no current layers $[\vec{H}, \vec{A}]$ is continuous over Σ , and, on the other hand, this same quantity is of the order of $1/r^3$ over S , so that the surface integrals disappear. We can therefore write

$$W_m^* = \frac{1}{8\pi} \int (\vec{H}, \text{rot. } \vec{A}) \delta v = \frac{1}{8\pi} \int (\vec{A}, \text{rot. } \vec{H}) \delta v.$$

As, according to Maxwell,

$$\text{rot. } \vec{H} = \frac{4\pi}{c} \left(\vec{J} + \frac{\vec{D}}{4\pi} \right),$$

we obtain

$$W_m^* = \frac{1}{2c} \int \left(\vec{A}, \vec{J} + \frac{\vec{D}}{4\pi} \right) \delta v.$$

In this formula, \vec{A} represents the vector potential of the currents of conduction in the elements $\delta v'$ different from δv .

In view of the result we obtained in (18) we have, therefore,

$$W_m^* = \frac{\mu_0}{2c^2} \iint \left(\frac{\vec{J}'}{r}, \vec{J} + \frac{\vec{D}}{4\pi} \right) \delta v \delta v'$$

This expression being symmetrical in relation to the quantities marked or not marked with an accent, we can exchange them and, considering (13), write

$$W_m^* = \frac{1}{2c} \int (\vec{J}, \vec{A}^*) \delta v, \quad \dots \quad (20)$$

or, still considering (16),

$$W_m^* = W_m + \frac{\mu_0 \epsilon_0}{4c^2} \int (\vec{J}, \text{grad}_P \psi) \delta v, \quad . \quad . \quad . \quad (21)$$

where $-W_m$ represents F. Neumann's electro-dynamical potential of currents of conduction.

The second term that figures in (21) represents the contribution of the currents of displacement which circulate practically between the extremities of the conductors, which constitute, according to Maxwell's theory, the necessary prolongation of the currents of conduction. This term admits a transformation which enables us to give to W_m^* the final interesting form. In fact, the function ψ under the form given in (6) can be written, considering (1) and (2) :

$$\psi = \frac{1}{\epsilon_0} \int (\vec{J}', \text{grad}_Q r) \delta v', \quad . \quad . \quad . \quad (22)$$

or even

$$\psi = \frac{1}{\epsilon_0} \int |\vec{J}'| \frac{\partial r}{\partial s'} \delta v',$$

designating by $\partial s'$ a linear element of the currents. Hence results

$$(\text{grad}_P \psi) = \frac{1}{\epsilon_0} \int |\vec{J}'| \frac{\partial}{\partial s} \left(\frac{\partial r}{\partial s'} \right) \delta v',$$

designating by ∂s another linear element of the currents. By substitution in (21) the second term on the right becomes

$$\frac{\mu_0}{4c^2} \iint |\vec{J}| |\vec{J}'| \frac{\partial^2 r}{\partial s \partial s'} \delta v \delta v',$$

so that W_m^* takes the form

$$W_m^* = \frac{1}{2} \frac{\mu_0}{c^2} \iint |\vec{J}| |\vec{J}'| \left(\frac{\cos \omega}{r} + \frac{1}{2} \frac{\partial^2 r}{\partial s \partial s'} \right) \delta v \delta v', \quad . \quad . \quad (23)$$

designating by ω the angle between the elements ∂s and $\partial s'$, and by r their mutual distance.

We may call the function $-W_m^*$ Helmholtz's electro-dynamical potential because Helmholtz started from this function to build a theory which admits Maxwell's theory as a limiting case †.

It is evident in (23) that for almost uniform closed currents the additional term relative to each filament of current

$$\oint i' \frac{\partial^2 r}{\partial s \partial s'} \delta s'$$

can be disregarded, as $\partial r / \partial s$ is a single-valued and continuous function of its variables. For open currents the contribution of this term is

† H. Poincaré, 'Electricité et Optique'; A. O'Rahilly, 'Electromagnetics.'

smaller in proportion as the extremities of the currents are closer. For example, in the case of the discharge of a plane condenser with a small thickness of dielectric, the contribution is very small, and the magnetic energy has more or less the same value as in a closed current.

Resuming, we have the following proposition explicitly enunciated by Planck: in quasi-stationary processes, the magnetic energy of a system of unclosed currents i_α can still be represented by the formula

[illegible]

which is established for closed currents by submitting the definition of the coefficients of induction $L_{\alpha\beta}$ to an extension which is determined in an obvious manner by the formula (23).

In the particular case of two unique linear currents, the additional term of (23) takes the form

$$\frac{1}{2} \int_a^b i \, ds \int_{a'}^{b'} i' \frac{\partial^2 \gamma}{\partial s \partial s'} ds',$$

where (a, b) and (a', b') designate the extremities of the conductors. The integral relative to s' integrated by parts gives

$$\frac{1}{2} \left| i' \frac{\partial r}{\partial s} \right|_{\alpha'}^b - \frac{1}{2} \int \frac{\partial r}{\partial s} \cdot \frac{\partial i'}{\partial s'} \delta s'.$$

If i' is almost uniform, the second integral is negligible. The same conclusion is valid for the integral relative to s . The complementary term in (23) can therefore be written

$$\frac{1}{2}ii'(r_{bb'} - r_{ba'} + r_{aa'} - r_{ab'}).$$

The coefficient of mutual induction of the two open circuits takes the form

$$L_u = \frac{\mu_0}{c^2} \int \int \frac{\cos \omega}{r} \delta s \delta s' + \frac{1}{2} \frac{\mu_0}{c^2} (r_{bb'} - r_{ba'} + r_{aa'} - r_{ab'}). \quad (25)$$

If one of the circuits is closed ($a \equiv b$, for example), the complementary term disappears; if almost closed, all takes place (approximately) as if it was rigorously closed.

XLII. *Notices respecting New Books.*

Sir AMBROSE FLEMING: (1) *Mathematics for Engineers* (3rd edition, 1943). (2) *Physics for Engineers* (1941; Reprinted 1942). (3) *Chemistry for Engineers* (1943). (4) *Mechanics for Engineers* (1944). (London: George Newnes, Ltd. Price 7s. 6d. each volume.)

SIR AMBROSE FLEMING's long, honourable and honoured life has been a very full one, and of few men can it be said with more justice that none of his

working days has ever to be counted as wasted. He has left his mark on the science of two generations ; but besides being a great engineer, he is—and the emphasis lies on the present tense—a great teacher and an indefatigable writer. We who have witnessed the strides made in wireless telegraphy since those days in the middle eighteen nineties when the news of a message sent over a distance to be measured in furlongs caused something of a stir, have reason to be grateful to Sir Ambrose for the treatises on the theory and practice of wireless which were the authorities for students and teachers of our generation.

And now, the approach of the 95th anniversary of his birthday (for Sir Ambrose was born in 1849) sees the veteran author hard at work, preparing a new series of text-books for a new generation of students. He believes—and his belief is justified—that there is need and want for a series of handy and inexpensive text-books in subjects ancillary to engineering which shall give the student just what he needs to illuminate his studies in his main subject. The series of volumes under review represent the author's attempt to supply the deficiency.

The characteristic Fleming touches are still there ; clarity, an intensely practical outlook, no wasted words, and the old habit of beginning at the very beginning and leading the reader rapidly to the really difficult parts of the subject.

It is impossible, in a brief review, to do more than give a summary of the contents of the four volumes. Physics and Mechanics are complementary volumes and should be studied together. The volume on Mechanics opens with a chapter on fundamentals—and it is interesting to note that the author, true to the old T and T' tradition, adopts the Newtonian definition of mass—and goes on to chapters dealing with particle dynamics, oscillations, central orbits, machines, general principles of dynamics, liquids, gases, prime movers, electromechanics, civil engineering appliances, and mensuration. The Physics volume leads off with a section on units and proceeds to discuss matter and elasticity, energy transformations, electrostatics and electromagnetism, currents, electric measurements and instruments, thermionics and applications, electron optics, sound, alternating currents and atomic transformations.

The Chemistry volume is thoroughly practical and industrial in its outlook, and, after a brief account of the atomic theory, the periodic law and allied matters, is concerned with such topics as the chemistry of iron and steel, of cellulose, of explosives, of petroleum, of insulating materials and plastics, of nitrogen fixation and ends with a chapter on the liquefaction of gases.

The volume on Mathematics (now in its third edition) is something of a *tour de force*. Beginning with the simplest operations of algebra the student is led rapidly but surely through the elements of trigonometry, co-ordinate geometry, calculus and differential equations, to brief but thoroughly useful surveys of vector algebra, harmonic analysis and hyperbolic trigonometry. All stages of the discussions are illustrated, whenever it is possible, by applications of special interest to students of engineering.

Such expositions are necessarily highly condensed, and the student who desires to make the fullest and most profitable use of this series of text-books will be well advised to study them under the direction of a competent teacher, who will comment on and expand the various sections and who will, above all, see that his pupil works through a sufficient number of illustrative examples.

There are a few slips and misprints which will doubtless be corrected in the later editions which will certainly be called for.

Sir Ambrose, who is one of the few survivors of the band of students who attended Clerk Maxwell's lectures, is to be warmly congratulated on his efforts to inspire a later generation of students with something of the Maxwellian traditions.

A. F.

Magnetochemistry. By P. W. SELWOOD. [Pp. 287.] (New York: Interscience Publishers, Inc., 1944. Price \$5.00. London: Imperial Bookselling Co.)

THE author defines magnetochemistry as the application of magnetic susceptibilities and of closely related quantities to the solution of chemical problems. While this definition enables Professor Selwood to produce a book of a reasonable size, it also means that the book must appeal mainly to chemists and to those physicists who are interested either in the determination of susceptibilities in a narrow sense or in acquiring a broad outlook on the magnetic properties of matter.

Little space is spent on the discussion of experimental techniques, and the mathematical discussion of magnetic theory is reduced to the absolute minimum. Yet the reader will note with pleasure those little touches which make an account of experimental work so much more interesting to read; *e. g.*, the author's own experiences of the disturbing effects of dissolved air on the susceptibility of water, and of the effects of traces of ferromagnetic impurity. The author possesses the happy knack of being able to describe difficult concepts in very few words. The reviewer particularly appreciated his discussion of the significance of the parameter Δ which occurs in the Curie-Weiss equation for the susceptibility of a paramagnetic as a function of temperature, and his neat description of the phenomenon of paramagnetic relaxation.

The remarkable growth of our knowledge of molecular paramagnetism in recent years is made clear in a chapter of 26 pages on this subject alone, which is followed by another of 43 pages on complex compounds. An account of some interesting work on the magnetic properties of human blood finds a place in the latter chapter, where we meet with the new idea of magnetic titration, or the determination of the susceptibility and course of a reaction mixture during the addition of a reagent.

The book is excellently printed and illustrated. It gives a survey of the magnetic properties of all the elements, their compounds and alloys, and brief accounts of ferromagnetism and applied magnometric analysis. It is full of valuable references and ought to find a place in every scientific library.

L. F. BATES.

Tables of the Bessel Functions $J_0(z)$ and $J_1(z)$ for Complex Arguments. Prepared by the Mathematical Tables Project. [Pp. xlv+406.] (New York: Columbia University Press, 1943.) \$5.00.

IN this table, values of the real and imaginary parts of J_0 and J_1 are given, when the argument is $re^{i\theta}$. Each is given to 10 decimal places, which means that the imaginary part has in places to be correct to 14 significant figures, since it tends to infinity with increasing r when θ approaches $\pi/2$. The tables are arranged with the functions of the zero and first orders on the same page, a succession of pages being devoted to a single value of the phase angle while r ranges from 0 to 10 at intervals of 0.01. The interval for θ is 5° . The table is thus much closer than that of its only forerunner, the table by Hayashi, which covered $r=0$ to 8 at the same interval, but only gave 17 values of θ between 0 and $\pi/2$. Interpolation in the present table is easy for r but by no means simple for θ if accuracy is demanded, and it would seem that at some future date it will be worth while for some organization to produce a table, say to 7 decimals, for each 1° of angle and perhaps 0.1 in r .

Like other volumes prepared by the Mathematical Tables Project, this carries a most instructive introduction, which includes an account of the types

of investigation in which Bessel functions of complex argument present themselves. As many readers will know, the Project was set up under the Work Project Administration, as part of the National Recovery machinery in the U.S.A. That it did not have to cease its extremely valuable activities is, we understand, due in part to the willingness of the Columbia University Press to take over the task of publication of its tables. The precautions detailed in the book for ensuring its accuracy are as thorough as is usual in this series.

J. H. A.

Table of the Reciprocals of the Integers from 100,000 through 200,000. Prepared by the Mathematical Tables Project. [Pp. viii+204.] (New York : Columbia University Press, 1943.) \$4.00.

As stated in the introduction to this volume, the only tables of reciprocals of comparable accuracy are those by Oakes and by Cotsworth respectively, but they are so arranged as to involve differences of 250 to 1000, for which linear interpolation is insufficient. In the present seven-figure tables, which have been entirely re-calculated, there are no differences greater than 100. Tables of proportional parts to $1/10$ of a unit are given on every page, and entries in the main table which have been obtained by rounding upwards are distinguished by a minus sign following the last figure, so that risk of introducing an error when interpolating is reduced to a minimum. The printing is very well done, but the figures are just a little smaller than is desirable for really comfortable use over any length of time.

J. H. A.

[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]

XLIII. *Two Methods of Measuring Ultra High Frequency Electric Fields.*

By K. RACHEL MAKINSON, B.A.(Cantab.), A.Inst.P., Commonwealth Research Scholar in the University of Sydney (1940-41)*, and H. D. FRASER, B.Sc. (Sydney) †.

[Received January 19, 1944.]

1. *Summary.*

Two methods are described which were developed, in connection with work on high frequency discharges in gases, for the measurement or comparison of electric field strengths of the order of 10 volts/cm. at frequencies up to 500 Mc/s.

The "induced dipole" method depends on the measurement of the couple exerted by the field on a small metallic strip lying at an angle of 45° to the electric vector. The instrument was calibrated at 1.2 Mc/s.; the calibration is independent of frequency up to higher than 500 Mc/s. A similar instrument was used by Zouckermann⁽¹⁾ to measure potential differences at u.h.f.

The "dielectric thermometer" method depends on the measurement of the rate of expansion of a liquid dielectric heated by the field under examination; this rate is proportional to the square of the field strength. This instrument was used to compare fields at two fixed frequencies, using normal propyl alcohol of uncertain purity as the liquid dielectric. It was found, by comparison with measurements made with the induced dipole, that calibration by means of the values of the complex dielectric constant of *n*-propyl alcohol of unspecified purity measured by Abadie⁽²⁾ was correct within the limits of experimental error at 136 Mc/s., but incorrect at 510 Mc/s. The dielectric thermometer was therefore calibrated against the induced dipole. Since commencing the study of this method we have learnt that the use of an instrument similar to the dielectric thermometer for measuring electric field strengths has been suggested by Divilkowsky and Filippof⁽³⁾ and by Zouckermann⁽⁴⁾.

The induced dipole method appears to be easily applicable at frequencies as high as 500 Mc/s., where the reliability of the data used restricted the use of the dielectric thermometer to the measurement of relative values of field strength. In measuring greater field strengths, permitting the use of a shorter conducting strip, the method using the dipole could be extended to still higher frequencies; it might also

* Now Assistant-Lecturer and Demonstrator in Radio-Physics, University of Sydney.

† Communicated by Professor V. A. Bailey.

profitably be developed to use a conducting strip one half a wave-length long.

The induced dipole is more sensitive and quicker to use than the dielectric thermometer, but less portable.

2. *The Field to be Investigated.*

Measurements of electric field strength at frequencies of the order of 500 Mc/s. are usually made with a loop aerial and detector. The use of this method, since it measures primarily the magnetic flux linkage with the loop, depends on a knowledge of the relation between the electric and associated magnetic fields. In the field which it was necessary to measure, in connection with work on high frequency discharges in gases, no simple relation was known, the electric field of wave-length 59 cm. being produced between two parallel copper plates 4 cm. square and 4.6 cm. apart, approximately balanced about earth. The field was generated by a coupled pair of magnetrons.

Any method of measuring field strengths at these frequencies, when it involves the use of electron tubes, requires very careful consideration of circuit and components if large errors are to be avoided.

The two instruments described here were therefore developed to measure this electric field more directly.

3. *The Induced Dipole.*

(a) *Description of Instrument.*

This is illustrated in fig. 1, where the conducting strip is shown suspended in a protecting glass tube.

The copper vane A, suspended by a fine quartz fibre B from a torsion head C, had dimensions 2.0 cm. \times 6.5 mm. \times 0.1 mm., its breadth vertically being made fairly large so that no additional damping device need be used. Attached vertically below the vane were three small plane mirrors M, as shown in figs. 1*a* and 1*b*, no. 1 being parallel to A. The mirrors were mounted on a small copper former, the size of the mirrors and of each arm of this former being about 2 mm. horizontally by 3 mm. vertically. In order to protect the fibre when the instrument was not in use, the whole assembly could be lowered through the torsion head by loosening the screw E, until the vane and mirrors rested in the bakelite support D, the screw E then again being tightened.

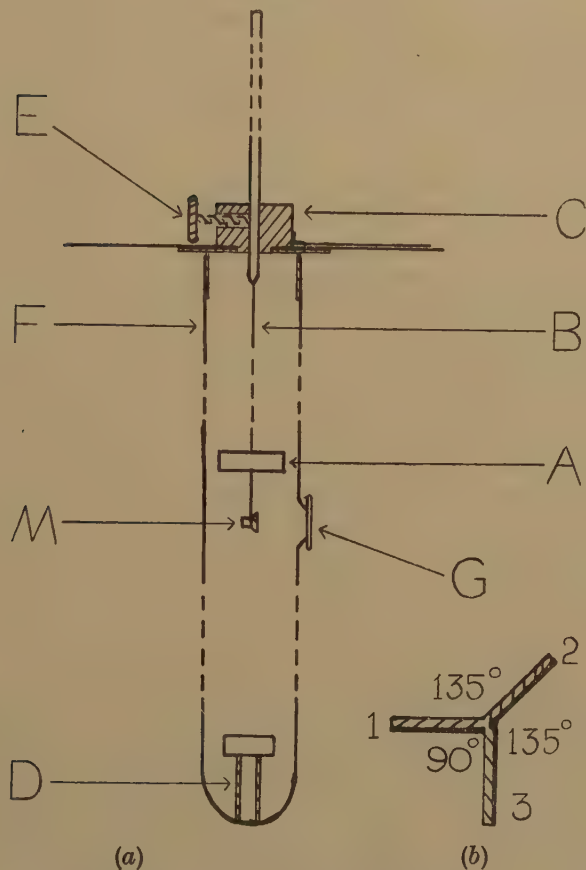
Before the apparatus was assembled the glass containing tube F and the thin cover glass G, which was cemented to F as a window, were warmed and coated with clear "Dulux" lacquer. This reduced to a negligible value screening of the copper strip from the external electric field by moisture deposited on the tube F (*v.* Appendix I). This screening was otherwise appreciable at the frequencies of a few hundred cycles per second used in an initial approximate calibration.

A small torque was found to be exerted on the strip by a strong magnetic as well as by an electric field. Calculations based on experi-

ments with a bar magnet of measured moment indicated that the torques due to the following fields would be negligible compared with the torque due to the electric field itself :—

- (i) the earth's magnetic field, and
- (ii) an alternating magnetic field of the order of that associated with the alternating electric field under examination.

Fig. 1.



(a). The induced dipole instrument.

(b). The arrangement of the mirrors M.

The reflecting surfaces are drawn heavily.

(b) Use of the Induced Dipole.

If a conducting strip of length L lies with its axis at an angle θ to a uniform electric field of intensity e , it experiences a torque due to the redistribution of its charges. This torque acts in the plane of L and the direction of the field and is proportional to e^2 , to L^2 and to $\sin 2\theta$. Provided that the resistance of the conductor, when enhanced by skin

to within 4 per cent., where ψ degrees was the rotation of the torsion head necessary to maintain the dipole at 45° to the field.

4. *The Dielectric Thermometer.*

(a) *Description of Instrument.*

The dielectric liquid was contained in a thin-walled, approximately spherical glass bulb with attached capillary tube for observation of the rate of expansion. The bulb was permanently screened from draughts by a thin paper cylinder, which was found to produce negligible screening of the electric field. The mean internal radius of the bulb was 0.73 cm., which is much less than the shorter wave-length (59.0 cm.) used. The mean thickness of the glass was of the order of 0.02 cm. The internal diameter of the capillary tube was of the order of 0.01 cm.

The sensitivities of various liquids were examined. Sodium chloride solution was found to be sensitive when sufficiently concentrated, but inconvenient to use owing to deposition of salt on the walls of the capillary tube. The sensitivities of methyl, ethyl and normal-propyl alcohols, glycerine, distilled water and 50 per cent. sulphuric acid were also examined. The *n*-propyl alcohol was found to be the most sensitive of these, in agreement with measurements of the dielectric constants of these liquids made by various authors. It was therefore decided to use *n*-propyl alcohol for the measurements.

(b) *Calibration of the Dielectric Thermometer.*

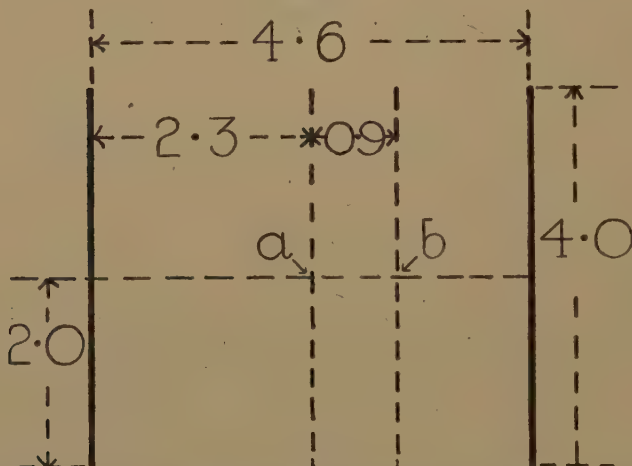
The complex dielectric constant of *n*-propyl alcohol has been measured as a function of frequency by Abadie⁽²⁾ and at some frequencies by Mizushima⁽⁴⁾. Abadie does not specify exactly the purity of the sample which he used, and his results differ considerably from those obtained by Mizushima. Our sample was about 95 per cent. pure. It is known⁽⁵⁾ that the dielectric constants of alcohols depend very markedly on the degree of purity, so that the dielectric thermometer could not be calibrated reliably from Abadie's values. The thermometer was therefore calibrated against the induced dipole at two frequencies, and this calibration was compared with that obtained from Abadie's values.

The frequencies used for the calibration were 136 ± 1 Mc/s. and 510 ± 7 Mc/s.

At both these frequencies it was found impracticable to maintain the field between plates which were large enough to permit the insertion of both instruments simultaneously into the same field. Measurements of a given field were therefore made with the induced dipole, then the dielectric thermometer was inserted in the field and the expansion-rate produced was noted, and finally the induced dipole was again used to check the constancy of the field. With the measurements at 136 Mc/s. this third set of observations was not always made, as the field was found to show no detectable variation over a period longer than the duration of the experiment.

The fields of frequency 510 Mc/s. which it was desired to measure in connection with work on ultra high frequency discharges in gases, as described in section 2, were maintained between two parallel copper plates which were 4.0 cm. square and 4.6 cm. apart. The 136 Mc/s. fields were maintained between the same plates. These fields were, therefore, not uniform over a distance comparable with the linear dimensions of either of the two instruments. The dielectric thermometer, being the smaller, was used to compare approximately the field strengths at two positions between the plates, being placed firstly centrally between them and secondly with its centre at 0.9 cm. from the plane midway between the plates on the line joining their centres (positions *a* and *b* respectively in fig. 2). The field strengths in position *b* were thus found, as shown in Table I. (section 4 (c)), to be about 12 per cent. greater than in position *a*.

Fig. 2.

Positions *a* and *b* of dielectric thermometer (lengths in cm.).

The measurements with the induced dipole were made with the instrument centrally between the plates at 45° to the line joining their centres (*v. fig. 3*), so that each end was 0.7 cm. from the midway plane.

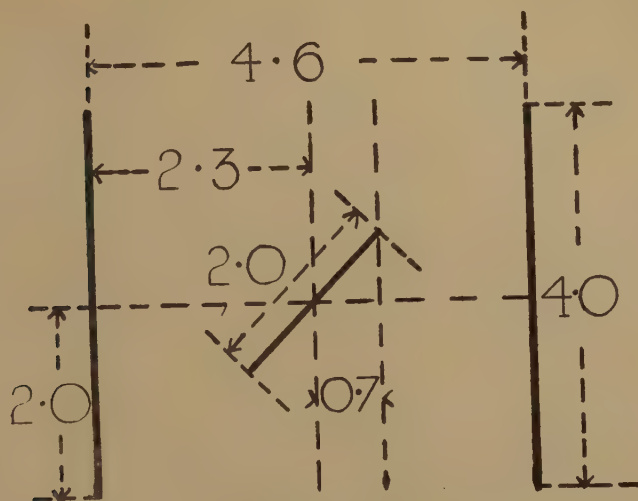
On the basis of a sketch of the lines of force between the plates, the field strength indicated by the dipole was expected to be somewhere between the actual field strengths at positions *a* and *b* and nearer the latter, so that the field in position *b*, a knowledge of which was required in order to calibrate the dielectric thermometer, was taken as equal to that indicated by the induced dipole, this value being possibly too low by less than 6 per cent.

(c) *Results of the Calibration of the Dielectric Thermometer.*

The results obtained are set out in Table I.

In this table *E* is the r.m.s. value of the field strength indicated by the induced dipole using equation 1 (section 3), and *r* is the rate of rise

Fig. 3.



Position of induced dipole instrument (lengths in cm.).

TABLE I.

Results of Calibration of Dielectric Thermometer against Induced Dipole.

Frequency (Mc/s.).	E (volts/cm.).	$r = dh/dt$ (cm./min.).		$(E_b - E_a)/E_b$ per cent.	$K = E/\sqrt{r_b}$ (volts min. ^{-1/2} cm. ^{-3/2}).
		r_a .	r_b .		
136±1	—	0.98	1.30	+13	—
	—	2.80	3.67	+13	—
	—	4.07	4.60	+ 6	—
	37.1	—	2.09	—	25.6
	37.4	—	1.87	—	27.3
	56.2	—	4.50	—	26.5
	59.6	3.85	4.56	+ 8	27.8
	61.4	—	4.91	—	27.7
					Mean 27.0±0.6
510±7	9.0	0.57	0.74	+12	10.5
	10.5	0.85	(1.09)	—	10.1
	12.5	1.05	1.37	+12	10.7
	12.8	1.39	2.06	+17	8.9
	18.2	2.30	3.14	+14	10.3
	19.6	2.43	3.04	+11	11.2
					Mean 10.3±0.5
Mean at both frequencies +12					

of the alcohol in the capillary tube of the thermometer. Measurements made with the dielectric thermometer in position *a* and in position *b* (see section 4 (*b*)) are denoted respectively by the subscripts *a* and *b*. The column $(E_b - E_a)/E_b$ gives the percentage difference between the fields at these two positions as measured by the dielectric thermometer.

As explained in section 4 (*b*) the field E_b at *b* is given by $E_b = cE$, where $1.00 < c < 1.06$.

Hence the calibration factor $K = E_b/\sqrt{r_b}$ of the thermometer may be taken as $E/\sqrt{r_b}$ within an accuracy of 6 per cent.

The value in brackets in column r_b is deduced from the corresponding value of r_a , using the observed average relation between r_a and r_b , which is

$$r_a = 0.78r_b.$$

The temperature of the room was not kept constant during these measurements; hence variations of *K*, from the mean value, up to 7 per cent. at 510 Mc/s., and up to 3 per cent. at 136 Mc/s., could be due to slow variations of temperature from day to day, with a consequent change in the dielectric constant and specific heat of the alcohol. The mean temperature during the experiments was about 18° C.

Remembering that an error of less than 4 per cent. is involved in the calibration of the induced dipole, we conclude that, at this temperature, the mean value of *K* at 510 Mc/s. is

$$10.3 \pm 0.9 \text{ volt min.}^{1/2} \text{ cm.}^{-3/2},$$

and at 136 Mc/s. is

$$27.0 \pm 1.7 \text{ volt min.}^{1/2} \text{ cm.}^{-2/2}.$$

Errors in the measurements of *E* and of $\sqrt{dh/dt}$ due to neglect of image forces are of the same magnitude, to within 1 per cent., so do not affect *K*.

The order of magnitude of the values of *K* obtained at 510 Mc/s. was confirmed to within 40 per cent. by Miss J. M. Freeman⁽⁶⁾, who, in the course of measurements in this laboratory on ultra-high frequency discharges in gases, calibrated the dielectric thermometer very approximately by the use of a symmetrical acorn diode voltmeter.

(d) *Theoretical Discussion of the Dielectric Thermometer.*

If the complex dielectric constant of the dielectric liquid at the frequencies concerned is known, the calibration factor at each frequency can be calculated. As explained in section 4 (*b*), values of the dielectric constant of *n*-propyl alcohol are available, but are of doubtful applicability to our specimen. It is of interest, however, to test their applicability by calculating the calibration factor using these data, and comparing it with the measured factor.

Let the dielectric constant of the alcohol be

$$\epsilon = \epsilon' - i\epsilon'' \text{ e.s.u., where } i = \sqrt{-1}.$$

If the electric field throughout the dielectric is uniform and equal to $e_1 = \Re E_1 e^{i\omega t}$, where \Re denotes the real part of the expression following it, then the mean rate, W , of absorption of energy per unit volume of the dielectric is given by

$$W = \frac{E_1^2 \omega}{8\pi} \epsilon'',$$

E_1 being supposed measured in e.s.u. and W in erg. sec.⁻¹ cm.⁻³.

We have now to relate the field e_1 inside the dielectric to the alternating field e_0 in its absence, since the induced dipole measures the field e_0 .

If the square of the radius of the bulb is very much less than the square of the wave-length the relation is formally the same as in an electrostatic field, with the complex value ($\epsilon' - i\epsilon''$) used for the dielectric constant. We then find for a sphere of dielectric, if the field e_0 is uniform for a distance very much greater than the radius of the bulb, that the field e_1 is not in phase with e_0 and that their amplitudes E_1 and E_0 respectively are related by the expression

$$E_1 = \frac{3E_0}{[(\epsilon' + 2)^2 + (\epsilon'')^2]^{1/2}}.$$

This expression is obtained by generalizing the expression given by Jeans ⁽⁷⁾ for the electrostatic case.

Hence

$$W = \frac{9\omega}{8\pi} \frac{\epsilon''}{(\epsilon' + 2)^2 + (\epsilon'')^2} E_0^2. \quad \dots \dots \dots (2)$$

The rate of rise, dh/dt , of the liquid in a capillary tube attached to the containing bulb is thus related to E_0 by the expression

$$E_0^2 = \frac{8\pi\alpha s J}{9} \frac{dT}{dh} \frac{1}{\omega\phi(\omega)} \frac{dh}{dt}, \quad \dots \dots \dots (3)$$

where

h = height of liquid in capillary (cm.).

t = time (sec.).

$J = 4.18 \times 10^7$ (erg cal.⁻¹).

T = temperature of dielectric (deg. C.).

ω = angular frequency of field e_0 (rad. sec.⁻¹).

$\phi(\omega) = \epsilon'' / \{(\epsilon' + 2)^2 + (\epsilon'')^2\}$.

α = density of dielectric (gm. cm.⁻³).

s = specific heat of dielectric (cal. gm.⁻¹ deg. C.⁻¹).

Errors may arise in applying this formula from the following sources :—

- (i) The assumption that e_1 and e_0 are related by the same formula as in the electrostatic case, *i. e.* that the bulb diameter is sufficiently small compared with a wave-length.
- (ii) Neglect of absorption of energy and distortion of the field by the glass.
- (iii) Forces due to surface tension and gravity in the capillary tube.

- (iv) Screening of the dielectric by a film of moisture of appreciable conductivity on the surface of the bulb.
- (v) The effect on ϵ_1 of the images of the bulb in the conducting plates bounding the field, the undisturbed field having been assumed uniform to infinity in deriving (3).

In these investigations the internal radius of the bulb was 0.73 cm. and the mean thickness of the glass of the order of 0.02 cm. At a wave-length of 59 cm. the error due to (i) was estimated to be less than 1 per cent. (see Appendix II.).

The conditions of the experiment were similar to those occurring in the investigations of Divilkowsky and Filippof⁽³⁾, who found that the absorption of energy by the glass was always considerably less than the absorption by the liquid, so that the liquid lost heat to the glass.

In our case the height of the liquid was plotted against the time and the curve obtained was consistently found to be linear for between 30 and 60 seconds after applying the field. The initial value of dh/dt was obtained from this curve, thus eliminating the correction for loss of heat from the liquid to the glass.

Calculation indicates that in our case the distortion of the field by the glass introduces an error into the relation between E_0 and E_1 , which is less than 2 per cent. at 136 Mc/s. and less than 1 per cent. at 510 Mc/s.

An estimate of the errors due to (iii) indicated that they were quite negligible, as also was the error due to (iv) at these frequencies.

When the bulb is at the centre of the field the error due to (v) is about 2 per cent., and when in the second position in which it was used (described later) about 3 per cent.

The total error in applying equation (3), due to the five causes above, is therefore less than 5 per cent. Hence we may expect the calculated calibration factor, $E/\sqrt{dh/dt}$, where E is the r.m.s. value of the field e , to be too high, by less than 5 per cent. We shall represent the calculated calibration factor by K' .

The quantity dh/dT was measured directly and was found to be 5.47 cm./°C., within 0.1 per cent., for normal-propyl alcohol contained in a bulb of internal diameter 1.45 cm. This value corresponds to expansion of the glass bulb as well as the alcohol, whereas, with heating by the electric field, initially only the expansion of the liquid is important; correcting for this we obtain $dh/dT=5.61$ cm./°C., which value has been used in the following calculations.

The density α was taken⁽⁸⁾ as 0.80 gm./c.c. to within 1 per cent. over the temperature range from 10° C. to 25° C.

The specific heat s was taken⁽⁹⁾ as 0.58 to within 5 per cent. over the same temperature range; at the mean temperature of our experiments, which lay between 15° and 21° C., it was within 2 per cent. of this value.

At a wave-length of 59.0 cm., for Abadie's⁽²⁾ sample of *n*-propyl alcohol

$$\epsilon'=9.67 \quad \text{and} \quad \epsilon''=7.97.$$

Using his figures, we obtain

$$\phi(\omega)=0.040\pm0.0012 \text{ at } 59.0\pm0.8 \text{ cm.}$$

Hence for the wave-length in use, which lay between these limits, the expression (2) becomes, on inserting the above figures,

$$E_0=10.7\sqrt{dh/dt}, \text{ or } E=7.6\sqrt{dh/dt}, \quad (4)$$

where E is the r.m.s. value of the field strength, dh/dt is now measured in cm./min. and E_0 and E in volts/cm.

The range of values inside which K' must lie, if Abadie's figures were appropriate to our case, is found, from the above discussion of the errors in the calculations, to be 7.0 to 7.8 volt min.^{1/2} cm.^{-3/2}.

At a wave-length of 220 ± 2 cm. (136 ± 1 Mc/s.), it is similarly found that, for Abadie's sample of alcohol,

$$\phi(\omega)=0.012\pm.0001.$$

If this value is appropriate to our sample, then

$$E_0=37.4\sqrt{dh/dt} \text{ and } E=26.5\sqrt{dh/dt},$$

the units being the same as in the equations (4).

TABLE II.

Comparison of K' and K . (volt min.^{1/2} cm.^{-3/2}).

Frequency (Mc/s.).	Calculated K' .	Limits of K' .	Mean K .	Limits of K .
136 ± 1	26.5	24.4, 27.3	27.0	25.3, 28.7
510 ± 7	7.6	7.0, 7.8	10.3	9.4, 11.2

The range of values inside which K' must lie is found to be 24.4 to 27.3 volt min.^{1/2} cm.^{-3/2}.

In Table II. these values of K' are compared with the values of K obtained by calibration against the induced dipole (section 4 (c)). The limits of K are deduced by assuming that the true value of K does not differ from the mean of the values given in Table I. by more than the sum of the probable errors in the measurement of K and in the calibration of the induced dipole.

The discussion in section 4 (c) as to the value of the field strength to be used in determining K , leads us to expect that $K'=cK$, where c is a constant and $1.00 < c < 1.06$.

It will be seen that there is no significant difference between the values of K and of K' at 136 Mc/s., but that the difference between them at 510 Mc/s. is three times the sum of the errors involved in the measurement of K' .

It appears that Abadie's values of the complex dielectric constant are

applicable to our sample of alcohol at 136 Mc/s., within the accuracy of our work, but not at 510 Mc/s.

We conclude that the dielectric thermometer cannot be used as an absolute instrument unless the sample of dielectric liquid is known to be of closely the same composition as a sample for which values of the dielectric constant at the frequencies concerned are available or can be determined. The dielectric thermometer is, however, useful for the comparison of fields at a fixed frequency, and may conveniently be calibrated against the less robust induced dipole instrument.

Our thanks are due to Professor V. A. Bailey, who suggested these methods of measuring field strengths, for his constant interest in the work, to Dr. R. E. B. Makinson for his assistance with the theoretical investigations of the appendices, and to Miss J. M. Freeman, who, in addition to making the measurements referred to in section 4 (c), assisted with the preliminary observations.

APPENDIX I.

Conditions Necessary for Calibration of Induced Dipole Field Strength Meter to be Independent of Frequency.

The two principal necessary conditions are as follows:—

(i) The length L of the dipole should be very much less than the wave-length λ , so that the field under investigation is approximately in phase over the length of the dipole.

In our case $L=2.0$ cm., $\lambda=59.0$ cm., so that $L/\lambda=0.03$.

(ii) The conductivity of the dipole at the frequency concerned should be high. The calibration is dependent on the conductivity if the currents induced in the dipole by the external field E are dependent on the conductivity, which is the case if the potential difference per unit length of the dipole due to its resistance is comparable with E . By regarding the charge on the dipole as concentrated near its ends, this charge can be approximately evaluated in terms of E . The current due to the alternations of this charge, and hence the p.d. per unit length of the dipole due to resistance, can then be estimated.

If R =the resistance per unit length of the dipole

and f =frequency (cycles/sec.),

it can be shown that the p.d. per unit length due to resistance is very much less than E if

$$R \ll 9/L^2 f \text{ e.s.u. cm.}^{-1}.$$

At 50 c/s. this condition becomes

$$R \ll 0.5 \cdot 10^{11} \Omega \cdot \text{cm.}^{-1},$$

and at 500 Mc/s.

$$R \ll 0.5 \cdot 10^4 \Omega \cdot \text{cm.}^{-1}.$$

Calculation of the resistance due to skin effect at 500 Mc/s. indicated that this second condition was satisfied.

The presence of a film of moisture on the dipole, by changing the effective value of R , could affect the calibration at low frequencies. To examine this we consider the screening effect of this film.

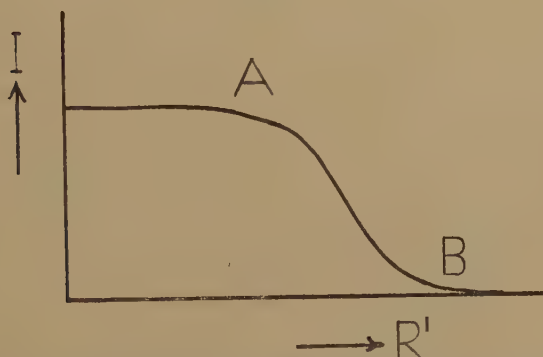
Let the resistance per unit length of the film of moisture be R' .

The metal will be screened from the external field by induced currents flowing in the film, but this screening will not be appreciable if the currents in the film are sufficiently small, which will occur if R' is high. Fig. 4 illustrates the general character of the variation with R' of the magnitude, I , of these induced currents for a particular frequency.

As in the case of the metal dipole, the portion AB of the curve, for which I depends markedly on R' , lies in the region for which

$$R' \sim 9/L^2 f \text{ e.s.u. cm.}^{-1}.$$

Fig. 4.



Variation of magnitude of induced currents in film of moisture with resistance per unit length of the film.

It can be seen that the necessary condition for currents in the film to be small is that

$$R' \gg 9/L^2 f \text{ e.s.u. cm.}^{-1}.$$

As before, at 50 c/s. this becomes

$$R' \gg 0.5 \cdot 10^{11} \Omega \cdot \text{cm.}^{-1}.$$

and at 500 Mc/s.

$$R' \gg 0.5 \cdot 10^4 \text{ ohm cm.}^{-1}.$$

Some information is available as to the resistivity of films of moisture deposited on glass under various atmospheric conditions⁽¹⁰⁾.

Assuming that the resistivity of films on metals is of the same order, we conclude that R' is unlikely to be much less than $10^8 \text{ ohm cm.}^{-1}$ ($10^{-4} \text{ e.s.u. cm.}^{-1}$) on days of normal humidity.

The frequency for which

$$9/L^2 f = 10^{-4} \text{ e.s.u. cm.}^{-1} \text{ is about } 20,000 \text{ c/s.}$$

The torque exerted on the dipole by a given field can, therefore, be

regarded as independent of the frequency and of atmospheric conditions provided that the frequency is very much greater than 20,000 c/s.

It is of interest to notice that, if $R' \ll 9/L^2 f$, the torque on the dipole would again be independent of frequency. The torque is only dependent on the frequency when $R' \sim 9/L^2 f$.

The film of moisture will also screen the metal dipole on account of its dielectric properties. An approximate calculation indicates that the screening from this source is negligible.

An initial rough calibration was made at 15,000 c/s., and the final calibration at about 1.2 Mc/s. ($\lambda = 250$ m.)

The two calibration factors ($E/\sqrt{\psi}$) differed by about 2 per cent., indicating that the range of frequency over which the torque exerted by the field is independent of frequency is wider than would be indicated by our approximate calculations.

Two other factors had also to be considered. For the torque produced by the field on the dipole system to be independent of the dielectric of the mirrors, it is necessary that the square of the length of each should be much less than L^2 . Since L was 2.0 cm. and the length of each mirror was 0.2 cm., this condition was satisfied.

It is also necessary that the dipole should not be screened from the external field by moisture on the protecting glass tube. This is the case if the resistance R_t , measured between diametrically opposite lines round that portion of the tube in the field, satisfies

$$R \gg 1/Lf \text{ e.s.u.}$$

This relation is obtained by considerations similar to those leading to the maximum permissible value for R .

As stated earlier, the tube was warmed and coated with clear Dulux lacquer, which is highly non-hygroscopic. R_t was measured and found to be $\sim 10^{12} \Omega$, thus satisfying condition (6) at frequencies above 50 c/s. Lest moisture was still present beneath the Dulux coating the screening effect of the tube on an electroscope was investigated at 50 c/s. and also with a constant voltage and found to be negligible.

The resistance R_t was measured again a year later and found to be unchanged in order of magnitude.

APPENDIX II.

Estimation of the Error in Treating the Problem of Calculating the Field Inside the Dielectric Thermometer as an Electrostatic Problem.

The power absorbed by the sphere from a plane progressive wave falling on it may be calculated by the method given by Stratton⁽¹¹⁾. The units used in this treatment are M.K.S. and $\epsilon = \epsilon' - i\epsilon''$ is the ratio of the complex dielectric constant of the dielectric composing the sphere to the dielectric constant of air. ϵ is then numerically the same as the value of the dielectric constant in e.s. units. The constants ϵ_2 and μ_2 of air have been taken as equal to those of free space.

Notation.

ϵ_2 =dielectric constant of air.

μ_2 =permeability of air.

$\epsilon=\epsilon'-i\epsilon''$ =ratio of the dielectric constant of the material of sphere to ϵ_2 .

μ_1 =permeability of the material of the sphere.

$N^2=\epsilon$.

ω =frequency of field.

$k_2=\omega\sqrt{\epsilon_2\mu_2}$.

a =radius of sphere.

$\rho=k_2a$.

E_0 =amplitude of undisturbed field.

W =rate of absorption of power by sphere.

All quantities are in M.K.S. units.

Calculation.

It is shown by Stratton⁽¹¹⁾ that

$$W = \frac{\pi E_0^2}{k_2^2} \sqrt{\frac{\epsilon_2}{\mu_2}} \left\{ \Re \sum_{n=1}^{\infty} (2n+1)(a_n^r + b_n^r) - \sum_{n=1}^{\infty} (2n+1)(|a_n^r|^2 + |b_n^r|^2) \right\},$$

where \Re means "the real part of",

$$a_n^r = - \frac{\mu_1 j_n(N\rho) [\rho j_n(\rho)]' - \mu_2 j_n(\rho) [N\rho j_n(N\rho)]'}{\mu_1 j_n(N\rho) [\rho h_n^{(1)}(\rho)]' - \mu_2 h_n^{(1)}(\rho) [N\rho j_n(N\rho)]'}$$

and

$$b_n^r = - \frac{\mu_1 j_n(\rho) [N\rho j_n(N\rho)]' - \mu_2 N^2 j_n(N\rho) [\rho j_n(\rho)]'}{\mu_1 h_n^{(1)}(\rho) [N\rho j_n(N\rho)]' - \mu_2 N^2 j_n(N\rho) [\rho h_n^{(1)}(\rho)]'}.$$

In these expressions the prime denotes differentiation with respect to the variable ρ or $N\rho$,

$$j_n(\rho) = \sqrt{\frac{\pi}{2\rho}} J_{n+\frac{1}{2}}(\rho),$$

$$h_n(\rho) = \sqrt{\frac{\pi}{2\rho}} N_{n+\frac{1}{2}}(\rho),$$

$$h_n^{(1)}(\rho) = J_n(\rho) + i n_n(\rho),$$

and $J_{n+\frac{1}{2}}$ and $N_{n+\frac{1}{2}}$ are the Bessel functions of the first and second kind respectively of order $(n+\frac{1}{2})$.

It is also shown by Stratton⁽¹²⁾ that, when n is integral,

$$j_n(\rho) = 2^n \rho^n \sum_{m=0}^{\infty} \frac{(-1)^m (n+m)!}{(m)! (2n+2m+1)!} \rho^{2m}$$

and

$$n_n(\rho) = - \frac{1}{2^n \rho^{n+1}} \sum_{m=0}^{\infty} \frac{(2n-2m)!}{(m)! (n-m)!} \rho^{2m}.$$

Hence, making the approximations $\mu_1 \doteq \mu_2$ and $\rho \ll 1$, we obtain

$$\begin{aligned} a_1^r &= + \frac{i}{45} (N^2 - 1) \rho^5 + 0(\rho^7), \\ *b_1^r &= + \frac{2i}{3} \left[\frac{N^2 - 1}{N^2 + 2} \rho^3 - \frac{1}{10} \frac{N^4 - 1}{N^2 + 2} \rho^5 + \frac{(10 + N^2)(1 - N^2)^2}{10(N^2 + 2)^2} \rho^5 \right] + 0(\rho^7), \\ b_2^r &= + \frac{i}{15} \frac{(1N^2 - 1)}{(2N^2 + 3)} \rho^5 + 0(\rho^7), \end{aligned}$$

where $i = \sqrt{-1}$.

All three of the above series converge rapidly when $\rho \ll 1$. Remembering that

$$N^2 = \epsilon = \epsilon' - i\epsilon'',$$

we obtain finally

$$\begin{aligned} W = \frac{\pi E_0^2}{k_2^2} \sqrt{\frac{\epsilon_2}{\mu_2}} \epsilon'' \rho^3 &\left[\frac{6}{q} + \rho^2 \left\{ \frac{1}{15} - \frac{\epsilon'^2 + \epsilon''^2 + 1 + 4\epsilon'}{5q} \right. \right. \\ &\left. \left. - \frac{1}{\epsilon''} \mathcal{I} \frac{(10 + \epsilon)(1 - \epsilon)^2}{(\epsilon + 2)^2} + \frac{5}{3} \frac{1}{[(2\epsilon' + 3)^2 + 4\epsilon''^2]} \right\} \right], \end{aligned}$$

where

$$q = (\epsilon' + 2)^2 + \epsilon''^2$$

and \mathcal{I} means "the imaginary part of."

When $\lambda = 59.0$ cm., we have $\epsilon' = 9.67$, $\epsilon'' = 7.97$ and $\rho = 0.074$.

$$\text{Hence} \quad W = C \rho^3 (0.239 + 0.18 \rho^2)$$

where

$$C = \text{a constant.}$$

The term in ρ^3 , which is proportional to ϵ''/q , is that which we obtained, in other units, by the electrostatic treatment (equation 2, section (4d)).

The correction to be applied to our calculated value of W in a given field is therefore

$$0.18 \rho^2 / 0.239 \quad \text{or} \quad 0.4 \text{ per cent.}$$

The correction to be applied to a value of the r.m.s. field strength E calculated from a measured value of W or dh/dt by means of equations 2 or 3, will then be 0.2 per cent., since $W \propto E^2$.

Our problem corresponds more nearly to the case of a plane stationary wave, but it is reasonable to assume that the correction is of the same order of magnitude as in the case of a plane progressive wave. On this basis we conclude that the error in our calculation of the calibration factor $K' = E/\sqrt{dh/dt}$, due to our electrostatic treatment of the problem is less than 1 per cent.

Note by Professor V. A. Bailey.

The investigation described in the foregoing paper was commenced by Mrs. Makinson in 1939 and has been delayed by the war.

* Stratton obtains the opposite sign for b_1^r and b_2^r and omits the second term in ρ^5 in the expression for b_1^r .

In spite of this it is desirable to publish the results obtained by her and Mr. Fraser as they may be of interest to physicists engaged in the study of h.f. fields.

References.

- (1) Zouckermann, *Annales de Physique*, xiii. p. 78 (1940), especially pp. 98-101. "Recherches sur la decharge en haute frequence avec electrodes exterieures."
- (2) Abadie, 'L'onde electrique,' xvi. p. 247 (1937), especially p. 257. "Recherches sur la dispersion anormale des liquide polaires. Methodes et applications."
- (3) Divilkowsky et Filippof, *Phys. Zeit. der Sowjetunion*, viii. p. 311 (1935). "Mesure des parties dielectriques a haute frequence dans les liquides."
- (4) Mizushima, 'Scientific Papers of Inst. of Physical and Chemical Research of Tokyo,' ix. p. 209 (1928). "Anomalous Dispersion and Absorption of Electric Waves."
- (5) Graetz, 'Handbuch der Elektrizitat und des Magnetismus,' i. p. 230 (1918).
- (6) Freeman, Joan M., 'Study of Ultra High Frequency Discharges in Gases at Low Pressures in the Presence of a Constant Magnetic Field.' Thesis for M.Sc. Degree presented to the University of Sydney. 1943.
- (7) Jeans, 'Electricity and Magnetism,' 5th ed. p. 228.
- (8) Thorpe, 'Dictionary of Applied Chemistry,' p. 397; and Hodgman and Holmes, 'Handbook of Chemistry and Physics,' 24th ed. p. 1006.
- (9) Hodgman and Holmes, *loc. cit.* p. 1872.
- (10) Hodgman and Large, 'Handbook of Chemistry and Physics,' 25th ed. p. 1864.
- (11) Stratton, 'Electromagnetic Theory,' pp. 563 *et seq.*
- (12) Stratton, *ibid.* p. 405.

XLIV. The Reception of Ohm's Electrical Researches by his Contemporaries.

By H. J. J. WINTER, M.Sc., A.Inst.P.,
University College of the South-West of England, Exeter*.

[Received March 15, 1944.]

G. S. OHM's experimental researches suffered neglect, in company with other such investigations in Germany, due to the widespread faith in Hegelism (*Naturphilosophie*). Hegel (1770-1831), who had less respect for the achievements of science than had Kant, regarded both Mind and Matter as essentials in the revelation of the Absolute, but maintained that Mind alone could afford knowledge of the Divine Plan independently of the methods of trial and error resorted to in the laboratory. With the later Hegelians the break between philosophy and science became

* Communicated by Professor F. H. Newman, D.Sc.

more apparent, so that the researches of an obscure teacher of mathematics at the Jesuit Gymnasium in Cologne could hardly be expected to make much impression upon the German universities⁽¹⁾, steeped as they were in the early nineteenth century in a strong classical and philosophical tradition⁽²⁾. It was not that individuals attacked Ohm personally (except G. F. Pohl), but that they were not at all interested in new experiments: nevertheless, these metaphysicians were certainly jealous for their own established position.

Amongst scientists themselves there was a certain reserve towards the application of mathematical methods of analysis to physical problems⁽³⁾. "Eminent students of science lived outside of the universities, belonging wholly or largely to the international republic which had its centre in Paris, exerting little influence on higher German education through the universities, and hardly any on German literature, which had meanwhile ripened into the age of classicism. This scattered condition of German science gave it on the one side a character which was foreign to the general tendencies of German thought, since this had come under the excessive influence of the speculative spirit without that wholesome check which exact research has always exerted⁽⁴⁾."

As a result there came into existence scientific journals as a medium of exchange of ideas between the scattered scientists, and to some of these Ohm contributed. He was greatly helped by his younger brother Martin Ohm in placing his MSS⁽⁵⁾.

The reception of Ohm's Law was also rendered more difficult within the purely scientific sphere due to

(1) The fact that, whilst Ohm had used a constant source of electromotive force, the majority of his contemporaries continued to experiment with batteries, with consequent confusion over questions of variable electromotive force due to change of concentration and to polarization; since neither of these factors was understood. Thus attempts to prove the law were unsuccessful, and gave rise to scepticism.

(2) Numerous unco-ordinated experimental data expressed in ill-defined terms of intensity and quantity⁽⁶⁾.

(3) The heated controversy between the rival hypotheses of the voltaic cell: the relative merits of Volta's contact hypothesis and the chemical hypothesis of Davy and Wollaston engaged the efforts of most galvanists. Supporting the former hypothesis were Ohm, Pfaff, Fechner, Marianini, and Jacobi, whilst Faraday, Roget, and De la Rive were in support of the latter.

On 27th July, 1826, Professor Erman, to whom Ohm had submitted his early work, made a report to the Prussian administration⁽⁷⁾. The main points were:—

- (i.) Dr. Ohm's researches have reduced many different electrostatic phenomena to two very simple formulae for the potential.

- (ii.) These formulæ are both new and adequate, but the first is comparable with that of Davy, and the second is outside existing analogies and could not be proved anyway.
- (iii.) It is doubtful whether Ohm's theory of the external circuit is valid for the internal charging mechanism of the voltaic pile itself.
- (iv.) The question still remains as to why conductors of frictional electricity do not exhibit just the same properties as those of the galvanic circuit.
- (v.) Ohm has not so far given the most essential part of his treatment, namely, an exact deduction of his two fundamental formulæ. All depends on this.

It must be remembered that 'Die galvanische Kette mathematisch bearbeitet' was to appear in the following year, and that the work submitted to Erman was tentative. In connection with point (iii.) Erman's researches⁽⁸⁾ had already helped in the explanation of the electrostatic phenomena of the voltaic pile.

When the full theoretical treatment did appear, copies were sent by Ohm to the Paris Academy of Science⁽⁹⁾ and to the Prussian Ministry of Education in Berlin. The latter copy⁽¹⁰⁾ was accompanied by a letter asking for resignation from his teaching post at Cologne and upholding experimental methods in science. In reply to a later communication, Schultze, the Minister of Education, a philosopher with no interest in experimental work, could only advise Ohm to accept the offer of a few lessons at the Artillery School in Berlin previously extended to him⁽¹¹⁾. Ohm, however, refused to be diverted from the purpose of improving his academic status, even though his difficulties in the fight for recognition seemed to grow steadily greater.

Two papers were now printed, one by G. F. Pohl, destructive in its criticism⁽¹²⁾, and the other by Kämtz of Halle⁽¹³⁾, which was so reserved as to give no help. Ohm wrote again to Schultze saying that he would like to reply to these statements, but had no one to support him; "Pohl is well known to be arrogant and his blindness in despising my work is due only to his own attempt to restrain me. He is misguided by his own animosity and not led by truth. But why is Kämtz so reserved? This I fail to understand⁽¹⁴⁾." Pohl had written of 'Die galvanische Kette': "he who looks on the world with the eye of reverence must turn aside from this book as the result of an incurable delusion, whose sole effort is to detract from the dignity of nature," and of Ohm's theory, that it was a web of naked fancies and had no support even in the most superficial observation of facts. Unfortunately, too, "the majority of galvanists of the day seemed contented in the labyrinth in which they had involved themselves."⁽¹⁵⁾ Further, there was strong disinclination against controversies in recognized journals. Ohm wished to publish his arguments in 'Literatur-Zeitschrift,' but the editor, Brandes⁽¹⁶⁾, advised him not to do so in view of the strength of the Hegelians, Pohl

and Schultze. After failing to obtain Schweigger's permission to insert a reply to Pohl in Schweigger's 'Journal,' Ohm succeeded in having his say in Kastner's 'Archiv.'⁽¹⁷⁾

Gradually Ohm gathered support; firstly from Pfaff⁽¹⁸⁾, who was in controversy with Pohl over galvanic questions, and then, between 1828 and 1832, from Schweigger and Fechner. Schweigger gave advice on the presentation of papers for publication, and cited Brewster's 'Journal' as a model⁽¹⁹⁾. Ohm wrote to him about the dependence of current upon the length of conductor and the nature of liquid in the battery. Schweigger's reply indicated that the current depended upon the "tension" of the metals and upon the concentration of the liquid, and made reference to Norremberg's observations⁽²⁰⁾. In a postscript to this letter of 1828 Schweigger added that with copper and zinc plates in spring water the depth of immersion determined the current, but seemed content to say that "you cannot deduce anything from this." The factors of electromotive force and current were not correctly distinguished, and the idea of internal resistance was not at all clear.

Ohm's debt to Fechner was greater than that due to Schweigger. Ohm had studied the published experimental work of Fechner, had made a copy, and had submitted a theoretical treatment of parts of it to Schweigger for publication. The latter forwarded the treatment to Fechner so that he might see it before printing, and Fechner wrote to Ohm wishing to settle their differences privately before entry into the 'Journal'⁽²¹⁾.

"I confess that a number of facts in my treatment were already known through the experiments of Jäger (who made important observations relating to the galvanic battery)⁽²²⁾, and some of my statements are included in your formulæ. I should have pointed this out more clearly in my work but for the fact that I was concerned in stressing further new observations which are at variance with these already known facts, and the first of which is not covered at all, or perhaps only partly, by Jäger's experiments. According to my formulæ I deemed it necessary to place restrictions upon yours. Of the accuracy of my formulæ there can be no one more convinced than I myself in the extent of their value in instances where I have been able to reproduce them by an alternative method.

"The differences which we had better smooth out seem to be included especially in the following two points:—

- "1. If you were to apply your formulæ originally derived for the closed circuit now to the open circuit you would find that there would have to be a uniform distribution of charge throughout the whole circuit beginning from the connecting surface. As my experiments show, this is only the case when the leads of both terminal conductors exceed a certain value so far undetermined.

- "2. You state that the experiments on the connection of the two plates

of one combination to earth simultaneously may be deduced from the principles of the closed circuit whereas I myself consider the deduction from the principle of the open circuit.

“As to the first question you are doubting the accuracy of my experiments. I could wish for nothing better than that you had a sufficiently accurate electroscope so that you could verify them for yourself. . . .

“In connection with the second point, I admit that all the results can also be deduced from your theory and your formulæ for the closed circuit, but this does not exclude the possibility of an alternative method of representing the electroscopical phenomena of the circuit when earthed, without having to consider the necessity of the plates being connected at the same time.”

Fechner then adds his appreciation of Ohm's Law :—

“The discovery of your fundamental law for the current in the circuit stating that the strength of the same is related to the reciprocal of the total resistance of all parts of the circuit, and the far-reaching importance of the consequences deducible from this law, have been entirely proved by me, not only by direct confirmation of your experiments but also through my own researches on the relevant conditions. Thus I would not forgive myself if I were to withhold from the public in a work which I am about to complete the really marvellous coincidence of facts which have been won for science . . . and if the few words I have added to the treatment Biot sent to you might not have persuaded you of the recognition due to you for the foundation of the principal laws of the electric circuit, then that recognition will be established by the third volume of my work”⁽²³⁾.

Fechner ends on a friendly note of co-operation. A letter from Fechner to Schweigger⁽²⁴⁾ also expresses great consideration of Ohm's work. Part of it runs :

. . . . “I share your estimation of Ohm in every way, because, except for his theoretical method of deduction, which seems to me to be capable of greater simplification, his experiments in my opinion have elucidated the electric circuit by means of very general conditions. . . .”

In the meantime Ohm was extremely active in extending his treatment to include all galvanic phenomena, and wrote a further paper on the galvanometer⁽²⁵⁾, for Schweigger's ‘Journal,’ 1829. Schweigger also advised him to get into touch with General Hellwig, whose influence would be of value⁽²⁶⁾.

About the question of polarization Fechner and Ohm did not entirely agree. This factor was largely responsible for the tardy recognition accorded to Ohm's Law for voltaic circuits, since though most investigators accepted its validity for thermo-electric circuits (and did not even repeat Ohm's researches of 1826), they were sceptical when confronted by the variations incurred with batteries. Fechner, however, in a letter dated 8th December, 1829⁽²⁷⁾, implored Ohm to do the relevant experiments himself, and added : “I expect I shall be able to continue my experiments

next spring. For the moment, I have concluded them with a statement about the experimental proof of all the laws which are based on your theory. You will find this confirmed in the next issue of Schweigger's 'Journal'". Then he thanked Ohm for making corrections to points in his *Lehrbuch* ⁽²⁸⁾.

Ohm next submitted the manuscript of "The hydro-electric (voltaic) circuit: does it follow the laws prescribed to it by the theory or not? Question and answer," to Schweigger-Seidel of Halle, a collaborator in the production of Schweigger's 'Jahrbuch.' Schweigger-Seidel, in reply ⁽²⁹⁾, sympathized with Ohm in his disappointments, and added: "Your contributions to the 'Jahrbuch' are most successful and I urgently ask you to go on publishing such important communications from time to time. I regret to state that here in Germany nothing is appreciated until it is published abroad. . . ."

" discussions are not refused by the 'Jahrbuch' provided that they deal only with facts and not with remarks and arguments alien to scientific work, nor moreover conducted by bickering people.

" Nature is the voice of the Lord. That is why all individuality must be eliminated."

And finally these consoling words:

"Rest assured that behind the clouds and the aery mist formations the light of the truth ultimately penetrates and will dissolve them in a smiling glance."

A further paper of Ohm's relating to the nature of unipolar conduction ⁽³⁰⁾ was criticized by Berzelius ⁽³¹⁾. "The explanation of Ohm is very simple but quite inadequate. A theory which does not explain all the phenomena of unipolarity cannot be the correct one." Ohm had already in his earlier paper on the galvanometer opposed certain comments of the same scientist, and now wished to answer him again. The influence of Berzelius was considerable and not to be regarded lightly, for it governed the labours and opinions of many German scientists. Schweigger, however, consoled Ohm ⁽³²⁾ :—

"To your request to me about making a decision as to whether your observation on some unimportant and lightly uttered remarks of Berzelius should be published, I should say with complete conviction that Berzelius does not regard himself as a physicist, and it seems to me that he has no feeling for physics, so it would be doing him too great an honour by replying in detail to remarks of his which are not worth discussing at all. . . . Don't bother. In time you will meet calmer conditions, which I sincerely wish for you. You won't get worried about things like that"

Editors of scientific reviews were several times compelled to advise Ohm on self-restraint, *e. g.*, Crelle, editor of 'Journal für Mathematik.' ⁽³³⁾ Ohm also wrote from Berlin a strong letter on 30th March, 1829, to Ludwig I. of Bavaria, asking for patronage ⁽³⁴⁾. In it he refers to Hegelism:

" my scientific work is of wide interest and has already attracted public notice, and I am sorry to say that it has found at the moment only one opponent (Pohl) whose opinion is based upon the principles of Hegel. . . . "

"The last pronouncement of Pohl is the only one which is against me, but his motives are so little concealed that I am mentioning it more for the sake of completeness than from fear of any ill-effects of this criticism, to which I replied in Kastner's 'Archiv.' " (35).

The King's answer was submitted by special order to the Royal Academy of Bavaria on 28th April, 1830. Ohm had offered his services and hoped to obtain some post under the Academy which would permit research. A special committee dealing with mathematical and physical questions reviewed "*Die galvanische Kette mathematisch bearbeitet*" with favour. On the committee were Stahl and Siber, extracts from whose reports are now recorded (36) :

Stahl : "The originator of the treatment under review has already shown himself, before the publication of this work, to be a very clever experimenter in electricity by his papers in several physical journals. In the actual work he has tried mainly to explain the laws by which the relevant factors operate in the galvanic circuit through the medium of analytical mathematical formulæ, and admittedly in as systematic a manner as possible. He initiated simple principles based on experiment and has completed the reasoning by really clever hypotheses. The results in his records agree with experience in so far as comparison can at present be made.

"To prove them all by himself the writer (Ohm) had neither the financial nor experimental means. By stating these few facts it will be seen that he dealt with the problem in a sagacious manner indicative of deep contact with electricity in general and with galvanism in particular, and that he is possessed of considerable mathematical ability.

"Furthermore, I should not fail to remark that he is the first German analyst to open up this field."

Stahl proposed that if Ohm were an efficient lecturer he should be appointed to a professorship of mathematics.

Siber was not helpful : "I cannot make a positive statement that I myself regard every experiment of applying mathematics to some part of physics as very valuable, as Ohm does."

But "it must be admitted that an attempt such as this one always has its own value."

. . . . "I do not know Ohm, neither his character nor abilities."

A third report (37) by the botanist Schrank was not so harsh as that of Siber. The Chairman of the Board of the Academy finally approached the philosopher Schelling, who then communicated with the King. Schelling's final answer (38) offered no assistance to Ohm. Ohm poured out his grief to Schweigger (39) :

"The birth of '*Die galvanische Kette*' has caused me dreadful pains

and I was about to damn the hour of its conception because of the shallow-minded people at the Court, who were incapable of understanding the real feelings of its mother."

But he was able to console himself in the loyal support of Fechner : " Particularly was I struck by the experiment which I could not understand at all, and with which he (Fechner) prefaces his work, an experiment upon which his opinion is especially based. . . .

" Fechner's demonstrations were, therefore, nothing but an amazing puzzle and the exchange of letters which has started between us was, as you know, characterized by difficulty ; however, now all the obscurity has been removed in a manner which bestows the same honour to my as well as to Fechner's method of approach. This wonderful man has repeated all the experiments in dispute eagerly and conscientiously. . . .

" My theory has found in him alone, if I am not mistaken, a very gallant defender ; and I have found into the bargain an honest friend through your intermediation."

Ohm went on to express appreciation of the fact that Schweigger was prepared to support the results of his contributions to the ' *Jahrbuch*.'

Ohm also wrote to Hofrat Kastner⁽⁴⁰⁾, editor of ' *Archiv für die gesammte Naturlehre*,' to thank him for the interest he had shown, and offering two further papers. In this letter he spoke against "intended deception such as that of Pohl," and stated his intention :—" I have decided to launch out in open attack against these attempts to withhold the true interpretations of nature, since I am convinced that those people who do this are not really concerned in the pursuit of truth but in the display of their own vanity."

The greater part of the third volume of Fechner's translation of Biot's ' *Experimental Physics* ' was now based upon Ohm's work on the electric circuit, and the preface struck a favourable note⁽⁴¹⁾ :—

" In representing the manner in which the factors of the galvanic circuit are related quantitatively I have followed the theory of Ohm and have further, by my own experiments, proved it, so that the most essential consequences of this theory have been confirmed by fact. I was very anxious to extend any developments of this theory to associated phenomena. I am not afraid to state of this theory that by it alone the true meaning was first introduced into the mechanism of the galvanic circuit. This work deserves my statement, and I feel that I owe this recognition also to the author of the above relationship." Ohm stressed this recognition of his work in a further letter, asking for more accurate apparatus, to the King of Bavaria⁽⁴²⁾. This application, together with a further one of February 1831 for a post in the gymnasium at Oldenburg (in which latter application Ohm quoted the nine papers he had produced to date)⁽⁴³⁾, met with no success. Nor did his application to succeed Fischer as Professor of Physics in the University of Berlin.

Ohm was next advised by Professor H. B. Hermann, recently appointed to the University of Munich, to apply as successor to Stahl⁽⁴⁴⁾. Siber's

report to the Academic Senate of the University of Munich supported Ohm and stressed the importance of 'Die galvanische Kette mathematisch bearbeitet' ⁽⁴⁵⁾; and further, the letter sent by the Senate to Ludwig I. of Bavaria was in Ohm's favour ⁽⁴⁶⁾. As a result, he was given a position not at Munich as he had hoped, but at the Polytechnic School in Nürnberg. Ohm, though warned by Hermann against Scharrer, the Principal of the School—who was noted for his opposition to practical science—accepted the appointment, made a great success of his work, and ultimately replaced Scharrer. A letter dated 23rd July, 1833, from Ohm's cousin Hilpert ⁽⁴⁷⁾ reveals that the appointment was largely due to the support given by Professor von Staudt, and that though no physical apparatus was available the requisition of such was being discussed. Von Staudt had a colleague, Engelhart, at the Polytechnic School, who wrote to Ohm asking him to contrive to demand the apparatus in spite of Scharrer, for "... you are well known as an excellent scientist" ⁽⁴⁸⁾. Ohm's ability as a teacher was now fully recognized and his scientific attainments, though less slowly, were gaining more admirers.

Meanwhile in England nothing had been heard of Ohm's Law. The classic researches of Davy remained as a model, and Davy's successors were still primarily concerned in the subsidiary static aspect which had limited his own work ⁽⁴⁹⁾. The manuscripts of Cavendish lay unknown in the possession of the Duke of Devonshire, and were to remain unrevealed for a further half century ⁽⁵⁰⁾. It was generally held (in error) that "conducting power" was directly proportional to the diameter (D) of a wire and inversely proportional to the square root of its length (L). A noteworthy advance was made by S. H. Christie (1785–1865), who determined experimentally the correct relationship, conductivity $\propto \frac{D^2}{L}$,

and first derived the principle of the bridge method of comparing resistance ⁽⁵¹⁾: this latter achievement was acknowledged by Sir Charles Wheatstone, who published practical details of a laboratory instrument ten years later ⁽⁵²⁾. English electrical work in particular perpetuated the obscure terms of intensity and quantity.

Ohm's theory took no account of the energy of the current. It was Roget who distinguished between the quantity of electricity and the tension at which it is produced, and then defined the latter as depending on the "energy of chemical action" of the cell; but Roget confused the issue by mentioning the term intensity again ⁽⁵³⁾.

E. T. Whittaker, with reference to this energy, wrote ⁽⁵⁴⁾: "This fact plus Faraday's discovery that quantity depends on the amount of chemicals consumed, places the origin of voltaic activity beyond all question."

The term intensity continued, however, to be used to denote tension (e.g. by Faraday (1834) and de la Rive (1836)) and no doubt added confusion to the already complicated controversy between the contact and chemical hypotheses concerning the origin of voltaic activity.

It should be noted that the discovery of the important quantity equation

$$Q=Ct$$

by Faraday, and his own research on electromagnetism and those of Henry, were made without knowledge of the existence of Ohm's Law. It is known that in 1833 Henry enquired of Dr. Bache where the theory of Ohm might be found ⁽⁵⁵⁾.

The influence of the experimental method spread gradually from the French Academy into German institutions; the travels of von Humboldt, its great supporter, however, occurred during the strongest period of Naturphilosophie, and his absence then was unfortunate. Nevertheless, the scientific spirit had already established itself at Göttingen with the remarkable researches of Gauss and W. Weber. Hegel died in 1831, and as the second quarter of the 19th century unfolded science came to assume equality with philosophical studies in the German universities. Ohm's work finally attracted attention as a result of French influence. The erroneous notion had arisen that Ohm had only deduced his law mathematically from debatable hypotheses. Pouillet believed this when he independently discovered the law by experiment and recorded his results in papers published in 1831 and 1837 ⁽⁵⁶⁾. Claims to priority made in connection with the latter paper now drew attention to the much-neglected work of 1826. An alternative notion, again erroneous, was that Ohm had obtained the results experimentally but only after previous theoretical deduction ⁽⁵⁷⁾. By 1840, however, a number of experimenters had verified Ohm's Law and had used it with success in their calculations ⁽⁴⁸⁾, and comments in its favour appeared in various research papers ⁽⁵⁹⁾; some of these are mentioned by Ohm in the letter quoted below ⁽⁶⁰⁾. We may note also Erman, Steiner, and Dove, who in 1839 fought successfully to establish the little-known G. S. Ohm as a corresponding member of the Prussian Academy of Sciences. Ohm can be cited as one of the last victims—almost perhaps in the personal sense—of the Hegelian régime in German education ⁽⁶¹⁾.

Ohm's reputation was at last firmly established through English science, as is shown by the award to him of the Copley Medal by the Royal Society in 1841 ⁽⁶²⁾, and the generous acknowledgment by Sir Charles Wheatstone in the Bakerian Lecture delivered before the Society in 1843 ⁽⁶³⁾.

An important letter written by Ohm to Ludwig I. of Bavaria on 6th May, 1842, summarized Ohm's successes to date:

(1) Investigators in favour of his theory were Fechner, Pfaff, and Poggendorff (Germany), Jacobi and Lenz (Russia), Henry in England, Munk af Rosenschöld in Lund, and Vorrsselmann de Heer (Holland).

(2) His principal work ('Die galvanische Kette mathematisch bearbeitet') had been translated into English some two years before, and a few weeks after the appearance of this an extract had appeared in French in 'Archivs de l'Electricité' (supplement to 'Bibliothèque Universelle Gèneve').

(3) The work had been acknowledged in his treatment on thermo-electric and voltaic circuits by Pouillet (ten years afterwards in 1837) in the transactions of the Paris Academy. Pouillet had adopted a part of Ohm's theory and had used most of his technical terms. This reproduction had been a subject for sarcasm in certain German periodicals. Ohm had sent a copy of his work in a very poor cover to the University of Paris, but did not expect Professor Pouillet "to adorn himself with foreign feathers." Pouillet had omitted the most difficult part of Ohm's theory, perhaps because he did not know German sufficiently well, but Ohm welcomed the support of such a competent referee.

(4) He had been made a corresponding member of the Royal Prussian Academy of Sciences, and of the Royal Academy of Science at Turin.

(5) "Towards the end of last year on November 25th, I received a letter from the Vice-President of the Royal Society in London to the effect that this Society had awarded me the Copley Medal for my experiments on the galvanic circuit, and though I have not yet received the medal, you will find a reference to the award in Poggendorff's 'Annalen,' so the matter may be regarded as settled." Ohm mentioned the annual award and great importance of the Copley medal, and the bestowal of this honour upon such men as Berzelius (1837) and Gauss (1839) before him. "I feel that such a distinction is too great for a work which at that time (1827) was planned only in outline and which had then no complete experimental foundation." This first recognition had come from abroad; at home he had fought Hegelism, and hoped that in future his experimental work would have the respect in Germany which it deserved.

A long letter was written on 26th April, 1844, by Poggendorff to Ohm, which gives certain interesting observations made by the former ⁽⁶⁴⁾ :

"If a circuit of electromotive force a is overcome by another circuit of greater electromotive force A , and p is the polarization of the weaker circuit, then the electromotive force of the system is $A-(a+p)$; that is, a has a constant value, which it keeps whether it is the cause of a current, or whether it only opposes A thereby weakening the current due to the latter."

Poggendorff went on to say that he now doubted the existence of the so-called resistance of transition which he had once upheld. He attributed his conversion to Ohm's treatment, and his previous opposition to an imperfect understanding on his own part of polarization. Poggendorff felt he could not accept Lenz's hypothesis ⁽⁶⁵⁾ to the effect that polarization is (1) constant, (2) independent of current, and (3) of sudden application.

And finally: "Do you think it possible that the liquids used in the galvanic circuit should be without individual conductivities?" and to "this delicate question" the remark: "This seems to me to be a fact but I should like to restrain myself for the moment."

A letter of 9th February, 1845, from Professor Hermann to Dr. Martius⁽⁶⁶⁾, Secretary for the mathematical and physical section of the Bavarian Academy, was instrumental in giving Ohm ordinary membership, and henceforth his advice was frequently sought on scientific problems⁽⁶⁷⁾. He was upheld by Thiersch, General Curator at Munich, and came to work with Siber, who had once made reports on him. A timely Italian statement, however, failed to exert much influence⁽⁶⁸⁾. Finally, Ohm became Ordinary Professor of Physics in the University of Munich in October 1852, thus achieving his life's ambition. After his death technical achievements based upon his law multiplied, but the discoverer and the manner of his discovery suffered neglect and misinterpretation⁽⁶⁹⁾.

In conclusion, I wish to thank Professor F. H. Newman for his help and encouragement, Professor E. N. da C. Andrade and Professor C. H. Lees for constructive criticism upon an earlier occasion, and Dr. Douglas McKie for his kindly interest.

References.

- (1) Ohm was partly to blame in not *stressing* his experimental work of 1826. He gave his theoretical treatment far greater opportunity of notice.
- (2) For the influence of Hegel, Geheimrat Schultze, and Freiherr von Altenstein on German education, see Paulsen, Bk. 2, S.316 ff. See also Helmholtz, 'Popular Lectures on Scientific Subjects,' 1862.
- (3) The critic Kämtz, to whom the Ministry of Education in Berlin had submitted some of Ohm's early work for comment, could not even follow the mathematics. Indeed "Schweigger selbst weisz nicht recht, was eine mathematische Behandlung soll." (Letter of Martin Ohm, Autumn 1827.)
See, for the full account,
- (4) J. T. Merz, 'History of European Thought in the 19th Century.'
- (5) Martin, Ohm's younger brother (1792-1872) had influential connections in Berlin and used these to help him, but Georg Simon was always reserved and somewhat independent. Martin Ohm placed "Vorläufige Anzeige des Gesetzes. . . ." (Schweigger's Journ. lxiv. pp. 110-118; Pogg. Ann. iv. pp. 79-88) simultaneously with Poggendorff and Schweigger on 25th May, 1825, and, in the summer of the same year, let Schweigger have "Ueber Leitungsfähigkeit. . . ." (Schweigger's 'Journal,' lxiv. pp. 245-247) and "Ueber Electricitätsleiter" (Schweigger's 'Journal,' lxiv. pp. 370-373). The rôle of Martin Ohm as an intermediary at the time of Poggendorff's suggestion of the use of thermoelectricity which made the discovery of Ohm's Law possible, must be judged important.
- (6) *e. g.* Joseph Henry spoke of cells in series as an "intensity" battery, those in parallel as a "quantity" battery. 'American Journal of Science,' xix. pp. 404-405 (1831): 'Smithsonian Miscellaneous Collections,' xxx. Pt. 1, p. 37 (1887). See also the earlier work of Francis Ronalds, Phil. Mag. xlv. p. 261 (1815).
- (7) 'Urkundensammlung des Deutschen Museums München,' No. 659.
- (8) Erman. Gilbert's 'Annalen,' xiii. p. 414. See also the work of Ritter and of Jäger, Gilb. Ann. viii. pp. 205, 207, 456, x. p. 11; Ohm, Ges. Abhandl. p. 52. Further, in Ohm's preface to 'Die galvanische Kette': ". . . the difference between the electrical forces occurring at any two points of the circuit is equal to the sum of all the tensions located

between these two points, and consequently increases or decreases exactly in the same proportion as this sum. When, therefore, one of these points is earthed, the sum of all the tensions located between the two makes its appearance at the other point. . . ." This statement interprets all the electroscope experiments with open pile due to Ritter, Jäger, and Erman.

- (9) A reply from J. Fourier stated that Dulong and Ampère would report on the work. 'Urkundensammlung des Deutschen Museums München,' No. 662. It was felt that a French or Latin version would be helpful.
- (10) Collected cited, No. 661. See also Lamont. Denkrede . . . (Bavarian Academy, Munich), 1855 : "Soon after the publication of this book in 1827, Ohm presented himself at the Ministry of Education in Berlin, and there met with a reception so little appropriate to the whole-hearted and self-sacrificing devotion to science of which he was conscious, that he felt it impossible to remain any longer in the public service."
- (11) See letter of G. S. Ohm dated 27th July, 1827 (Berlin) to von Altenstein. 'Urkundensammlung des Deutschen Museums München,' No. 3351. R. von Füchtbauer wrote of Schultze "dass er für Ohm und sein Tun noch keinen massstab gefunden habe, indem er ihn fortwährend wie einen behandle, den alles an seiner physikalischen Existenz gelegen sein müsse" (George Simon Ohm, ein Forscher wächst aus seiner Väter Art, Berlin, 1939).
- (12) Berliner Jahrbüchern für wissenschaftliche Kritik. Nos. 11-14 (1828).
- (13) Allgemeinen Literaturzeitung, Leipzig, Nos. 13, 14 (1828).
- (14) Letter by G. S. Ohm (6th March, 1828 ; Berlin) to Prussian Ministry in Berlin. Collection cited, No. 669.
- (15) G. F. Pohl, *op. cit.*
- (16) Letter from H. W. Brandes (24th March, 1828 : Leipzig) to G. S. Ohm. Collection cited, No. 706.
- (17) Kastner's Archiv. Bd. 14, H. 4 (1828).
- (18) Letter from J. W. Pfaff (to G. S. Ohm) (10th Oct. 1827, Erlangen) Collection cited, No. 688. It is important to note that Kastner and Pfaff were the first to realize the value of Ohm's work. J. W. Pfaff, Professor of Mathematics at Erlangen, and an old friend of the brothers Ohm, died in 1835.
- (19) Letter from Schweigger to G. S. Ohm (18th Oct. 1828, Halle). Collection cited, No. 689.
- (20) The same letter, also its postscript.
- (21) Letter from G. T. Fechner to G. S. Ohm (14th Nov. 1828, Leipzig). Collection cited, No. 692. See also G. T. Fechner, Schweigger's Journ. liii. pp. 429-441 (1828).
- (22) Jäger. See ref. (8) above.
- (23) G. T. Fechner, 'Lehrbuch des Galvanismus und Elektrochemie' (Bearbeitung von Biot's 'Lehrbuch der Experimental-Physik.)' L. Blosz, Leipzig, 2nd ed. 1829.
- (24) Letter from Fechner to Schweigger (17th Nov. 1828, Leipzig). Coll. cit. No. 690.
- (25) G. S. Ohm, Schweigger's Journ. lv. pp. 1-64 (1829).
- (26) Letter from Schweigger to G. S. Ohm (7th April, 1829, Halle). Coll. cit. No. 698.
- (27) Coll. cit. No. 702.
- (28) *Op cit.* ref. (23).
- (29) Letter from Fr. W. Schweigger-Seidel to G. S. Ohm (21st April, 1830, Halle). Coll. cit. No. 703.
- (30) G. S. Ohm, Schweigger's Journ. lix. pp. 385-435 (1830) ; lx. pp. 32-59 (1830).

- (31) J. Berzelius, 'Jahresbericht über die Fortschritte der physischen Wissenschaften, aus dem Schwedischen übersetzt von F. Wöhler,' 11th year, page 21 (1832).
- (32) Letter from Schweigger to G. S. Ohm (22nd July, 1832, Halle). Coll. cit. No. 704).
- (33) Letter from Crelle to G. S. Ohm (15th May, 1829, Berlin). Coll. cit. No. 699.
- (34) 'Akten des Bayerischen Staatsministeriums für Unterricht und Kultus, München.'
- (35) See reference (17).
- (36) Records cited in ref. (7).
- (37) 'Akten der Bayerischen Akademie der Wissenschaften München.'
- (38) Letter from Schelling to Ludwig I. of Bavaria (10th May, 1829, Munich). Records cited in ref. (37).
- (39) (Written from Berlin) 'Urkundensammlung des Deutschen Museums München,' No. 686.
- (40) *Ibid.* No. 695.
- (41) See ref. (23).
- (42) Letter from G. S. Ohm to Ludwig I. (27th Nov. 1829, Berlin) 'Akten des Bayerischen Staatsministeriums für Unterricht und Kultus, München.'
- (43) Letter "an das groszherzoglich Oldenburgische Konsistorium." 'Urkundensammlung des Deutschen Museums München,' No. 707.
- (44) Letter, 15th Feb. 1833, Munich. *Ibid.* No. 724. For Ohm's application addressed to the Minister, von Wallenstein, on 23rd Feb. 1833, and signed "Lehrer der Mathematik an der allgemeinen Kriegsschule sowie an der vereinigten Artillerie—und Ingenieurschule zu Berlin," see 'Akten des Bayerischen Staatsministeriums für Unterricht und Kultus, München.' F. B. Hermann had been a student friend of Martin Ohm when at Erlangen.
- (45) Dated 4th April, 1838, Munich. *Ibid.*
- (46) Signed Dr. Oberndorfer, Rektor (18th April, 1833, Munich). *Ibid.*
- (47) Written from Hilpert at Nürnberg to G. S. Ohm in Berlin, 'Urkundensammlung des Deutschen Museums München,' No. 726.
- (48) Letter from F. Engelhart, 6th Aug. 1833, Nürnberg). *Ibid.* No. 728.
- (49) James Cumming, Trans. Camb. Phil. Soc. ii. p. 47 (1823). Snow Harris, Edin. Trans. 1831; Phil. Trans. 1832. S. H. Christie (below).
- (50) Inspected and arranged for publication by James Clerk Maxwell, who verified many of the observations. 'The Researches of the Hon. Henry Cavendish.' Cambridge, 1879.
- (51) S. Hunter Christie, Bakerian Lecture entitled "Experimental Determination of the Laws of Magneto-electric Induction," Phil. Trans. cxxiii. pp. 95, 142 (1833).
- (52) The "differential resistance measurer." Bakerian Lecture, 1843 (see H. J. J. Winter, Phil. Mag., October 1943).
- (53) Peter Mark Roget. "The absolute quantity of electricity which is thus developed, and made to circulate, will depend upon a variety of circumstances, such as the extent of the surfaces in chemical action, the facilities afforded to its transmission, etc. But its degree of intensity, or tension, as it is often termed, will be regulated by other causes, and more especially by the energy of the chemical action." 'Galvanism,' § 70 (1832).
- (54) E. T. Whittaker, 'A History of the Theories of the Æther and Electricity in the 19th Century.' Dublin.
- (55) Mary A. Henry, 'Electrical Engineer,' xiii. pp. 28–30. Henry had to await his own visit to London in 1837 before his query was answered.

Alexander D. Bache, grandson of Benjamin Franklin, visited Europe to study educational methods, 1836-1838.

- (56) Pouillet. The following two quotations reflect the tendencies in German philosophy which accorded pre-eminence to speculative thought. " (Mr. Ohm), n'a donné de la justesse de sa pensée qu'une démonstration mathématique, fondée sur des considérations d'électricité statique, qui, aujourd'hui encore auraient elles-mêmes besoin de démonstration. Mr. Ohm, en un mot, a donné cette loi, non pas comme conséquence—de principes avoués et reconnus, mais comme conséquence d'une pure hypothèse.—(Pouillet, C. r. xx. p. 209.)
- " Et puisque nous faisons une réserve, nous en profiterons pour protester hautement contre la tendance illogique qui règne dans ce mémoire (de Mr. Weber) dans le livre de Mr. Ohm, dont les Archives ont rendu compte, et dans d'autres productions des premiers savants de l'Allemagne, tendance qui consiste à présenter les résultats de recherches instrumentales comme les conséquences de certaines lois qu'on pose en prémisses et a priori, au lieu de montrer, conformément à la vérité historique, que ces lois découlent au contraire des expériences, qu'on a faites dans un but plus ou moins déterminé, et d'avouer, qu'elles sont entachées de toutes les incertitudes et de toutes les irrégularités, qui peuvent être inhérentes dans l'état actuel de la science expérimentale aux procédés d'observations qu'on a choisis.—(De la rive, Arch. de l'El. v. p. 449.)
- (57) This error, and others, still persist, *vide* Proc. Phys. Soc. xlviii. p. 562 (1936).
- (58) See E. Lenz, Pogg. Ann. xxxiv. pp. 418-437 (1835); M. H. Jacobi, Pogg. Ann. xlviii. pp. 26-58 (1839) in Russia; P. S. Munck af Rosenschöld, Pogg. Ann. xliii. pp. 193-228 (1838) in Lund; P. O. C. Vorsselman de Heer, Pogg. Ann. xlvi. pp. 513-537 (1839) in Holland; J. C. Poggendorff, Pogg. Ann. xlix. pp. 31-72 (1840) in Germany.
- (59) e. g. M. H. Jacobi, Professor in the University of Dorpat. 'Electromagnetic Experiments,' series iii. (in Taylor's Scient. Mem. ii. p. 8). "Considerably enlarged by the ingenious researches of M. Lenz and conjointly with the electro-chemical views of Mr. Faraday, this theory has become capable of connecting under one single point of view a multitude of facts."
- (60) 'Akten des Bayerischen Staatsministeriums für Unterricht und Kultus, München.'
- (61) H. von Treitschke, 'Deutsche Geschichte im 19 Jahrhundert.' 4 Teil, S. 469 (1889).
- (62) Proc. Roy. Soc. iv. p. 336 (1841). See also the Obituary Notice in Proc. Roy. Soc. vii. pp. 598-602 (1854-55).
- (63) Wheatstone also helped "Mr. William Francis, Student in Philosophy in the University of Berlin" in the first English translation of 'Die galvanische Kette mathematisch bearbeitet' (Richard Taylor's 'Scientific Memoirs,' vol. ii. pp. 401-506 (1841). This English version appeared six years after the last German copies had disappeared from the booksellers' shops in Berlin.
- (64) Written from Berlin. 'Urkundensammlung des Deutschen Museums München,' No. 735. See also J. C. Poggendorff, Pogg. Ann. lx. (1843); and lxiii. pp. 389-405 (1844).
- (65) E. Lenz, Pogg. Ann. lii. (1842); Lenz and Saweljew, Pogg. Ann. lxvii. p. 497 (1846).
- (66) Written from Munich. 'Akten der Bayerischen Akademie der Wissenschaften, München,' 1845, 6 Sitz. der 2. Klasse. Beil. 4.

- (67) 'Urkundensammlung des Deutschen Museums München,' Nos. 737-739. 'Akten der Bayerischen Akademie der Wissenschaften, München.'
- (68) Dr. Achilles Perugia, 'Teoria matematica del circuito galvanico,' Pisa, 1847. This contained a long appreciation of Ohm's work; *Confronto dei risultati teoretici ottenuti coi risultati dell' esperienza*. Such a supplement prefaced to the first edition of 'Die galvanische Kette' would, says Bauernfeind, have served a valuable purpose.
- (69) In 1860 a French version of 'Die galvanische Kette' due to Gaugain was published in Paris. A new edition in German, with an introduction by Dr. J. Moser appeared much later (Wien, 1887). Thus the theoretical work was given unwarranted precedence. Some German text-books mentioned Ohm's *theory only*: others omitted even his name, as did Dr. Jos. Krist, 'Anfangsgründe der Naturlehre,' bearbeitet von Dr. W. Pscheidl (Wein and Leipzig, 1893). An interesting passage occurred in 'Lehrbuch der Physik und Meteorologie' (Müller-Pfaundler) Wieweg, Braunschweig, 8th edn. 1881; Book 4, on p. 298:—

"Auch Pouillet machte Anspruch auf die Entdeckung des Gesetzes der Stromstärke, welches er auf experimentellem Wege aufgefunden, während Ohm, seine Resultate aus theoretischen Betrachtungen abgeleitet hat, auf die wir später zurückkommen werden. Wenn man auch zugiebt, was sehr leicht möglich ist, dass beide Gelehrte ganz unabhängig von einander und auf ganz verschiedenen Wegen dasselbe Gesetz fanden, so muss man doch unbedingt Ohm die Priorität zugestehen, indem die Publication seines Werkes älter ist, als die Pouillet'schen Versuche."

In England the significance of the *experimental* paper of 1826 was first explained by N. de V. Heathcote, 'Science Progress,' xxvi. pp. 51-75 (1931-32).

For more details relating to Ohm's correspondence, the reader is referred to Ohm's 'Handschriftlichem Nachlass' (L. Hartmann), München, 1927.

XLV. *A Note on Two Definitions of Noise Figure in Radio Receivers.*

By D. K. C. MACDONALD, M.A., A.Inst.P., Captain, R.E.M.E.,
Instructor, Military College of Science, Fire Control School*.

[Received April 13, 1944.]

1. *Introduction.*

THERE are known to the writer two definitions of noise figure (or factor of merit as regards signal to noise performance) of which one is in a fundamental form in general terms, most likely to appeal to a "radio physicist," the other in a form more suited to the ideas of the "radio engineer." The purpose of this note is to show that these lead to the same result, but that while the second definition is, in general, more suitable for immediate calculation with the data available to those

* Communicated by the Author.

engaged in radio engineering, yet care is necessary to avoid error in the computation of noise figure for receivers in which an amplifier stage (or stages) after the first is significant. In such cases, the more fundamental form is of value as a check on the work.

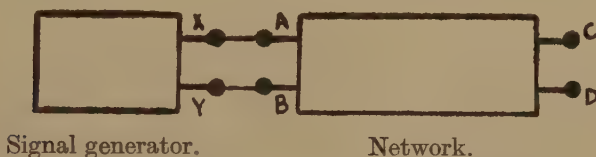
It should be appreciated that in this note we are only concerned with radio frequencies above the limit where external noise such as "atmospheric static," "ignition noise," etc., is significant. As a rough guide we may say that above about 70 megacycles/sec. external noise may be neglected.

2. Definitions of Noise Figure.

2.1. "Fundamental" Definition of Noise Figure.

In fig. 1 the equipment whose noise figure we wish to determine is represented by a generalized network with two input terminals A and B, and output terminals C and D. It should be appreciated, and becomes evident from examination of any particular case, that the noise figure of a network is meaningless unless the signal generator "driving" it is specified; that is, in general, the noise figure of the equipment is a

Fig. 1.



function of the output impedance of the signal generator. All "matching" or "coupling" elements are for convenience considered as an integral part of the generalized network.

2.11. Gain of Network.

Let S_i be "available" signal power⁽¹⁾ from signal generator at terminals X and Y.

Let N_i be "available" noise power⁽²⁾ from signal generator at terminals X and Y.

The signal generator is considered to generate only thermal noise and therefore:

$N_i = kTB$, where B = "noise" (integrated) band-width of the network when driven by specified signal generator.

k = Boltzmann's constant.

T = Effective temperature (°K.) of signal generator, generally accepted as room-temperature.

Let S_o be "available" signal power⁽¹⁾ from network at terminals C and D when driven by the signal generator.

Let N_o be "available" noise power⁽²⁾ from network at terminals C and D when driven by the signal generator.

thermal noise at room-temperature. The only remaining noise source is then R_g , the effective thermal resistance in the first grid circuit.

Thus :

$$F = \frac{e_a^2}{e_s^2} \cdot \frac{4(R_{eq} + R_g)kTB}{4R_a kTB}$$

$$= \frac{e_a^2}{e_s^2} \cdot \frac{(R_{eq} + R_g)}{R_a} \quad \dots \dots \dots (4)$$

3. Equivalence in a Simple Case.

Fig. 3 (a) represents a simple one-stage triode amplifier whose noise figure we wish to examine. We shall consider the case where the aerial

Fig. 3 (a).

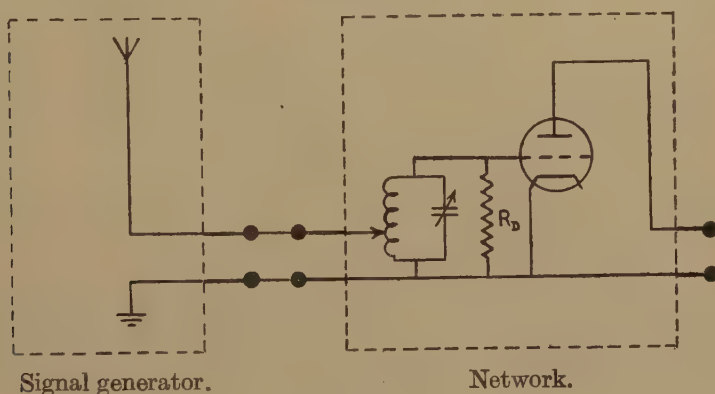
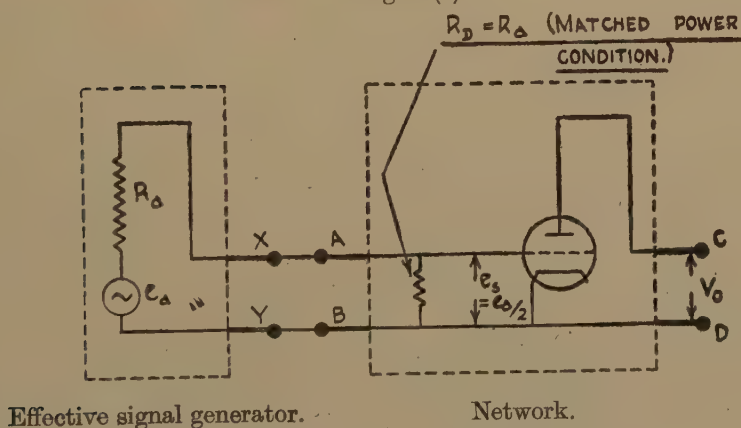


Fig. 3 (b).



R_a :—Transformed aerial radiation resistance.
 e_a :—Transformed aerial open circuit voltage.

is matched to the grid circuit for maximum power transfer. Fig. 3 (b) then represents the essential features as regards signal.

$$S_i = \frac{e_a^2}{4R_m}$$

$$S_0 = \frac{v_0^2}{4\hat{r}} = \frac{\mu^2 e_s^2}{4\hat{r}} = \frac{\mu^2 e_a^2}{16\hat{r}} \begin{cases} \hat{r} = \text{Anode a.c. resistance of valve.} \\ \mu = \text{Voltage amplification factor of valve.} \\ e_s = e_a/2. \end{cases}$$

$$G = \frac{S_0}{S_i} = \frac{\mu^2 R_a}{4\hat{r}}$$

$$N_i = kTB.$$

The output noise N_0 is composed of two parts—amplified thermal noise, $N(t)$, from the effective resistance in the grid circuit, and shot noise, $N(s)$, in the anode circuit. The effective thermal resistance in the grid circuit is $R_g = \frac{R_a}{2}$. By straightforward circuit analysis we find :

$$N(t) = \frac{\mu^2 R_g}{\hat{r}} \cdot kTB.$$

If i_n be the r.m.s. short-circuit anode shot noise current, then :

$$N(s) = \frac{i_n^2 \cdot \hat{r}}{4}.$$

Hence, from equation (2) :

$$F = \frac{e_a^2}{4R_a} \cdot \frac{1}{kTB} \cdot \left\{ \frac{\mu^2 R_g}{\hat{r}} \cdot kTB + \frac{i_n^2 \hat{r}}{4} \right\} \cdot \frac{16\hat{r}}{\mu^2 e_a^2} \dots \dots \dots (5)$$

To simplify this expression we use the definition of “equivalent grid noise resistance,” R_{eq} :

$$i_n^2 = 4kTBR_{eq}g_m^2. \quad (g_m = \text{mutual conductance of valve.})$$

Thus :

$$F = \frac{4}{R_a} \cdot (R_g + R_{eq}). \quad \dots \dots \dots (6 a)$$

(Simplifying from (5) and using $\mu = g_m \hat{r}$.)

Or :

$$F = 2 + \frac{4}{R_a} \cdot (R_{eq}). \quad \dots \dots \dots (6 b)$$

(Since $R_g = R_a/2$.)

Comparison of (6 a) derived from the “fundamental” definition shows immediate correspondence with (4), the “engineering” definition, remembering that in the example chosen $e_s = \frac{e_a}{2}$ because of the matched power condition.

This result can be generalized without difficulty to show equivalence of the two definitions for any arbitrary conditions in the input circuit of the single stage amplifier.

4. *Receivers with More than One Stage.*

4.1. “Fundamental” Derivation of Noise Figure.

Fig. 4 represents a receiver with an initial stage (network 1) and remaining stages (network 2), driven at the input by a specified signal generator,

Let F_1 be the noise figure of network 1 when driven by the specified signal generator.

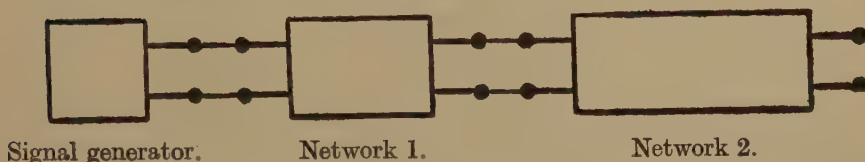
Let F_2 be the noise figure of network 2 when driven by a signal generator of output impedance equal to that of network 1.

Let G_1 be the gain of network 1 as defined in 2.11.

Then we can derive the general result that the noise figure of the complete receiver when driven by the specified signal generator at the input is given by :

$$F_{1+2} = F_1 + \frac{F_2 - 1}{G_1} \quad \dots \quad (7)$$

Fig. 4.

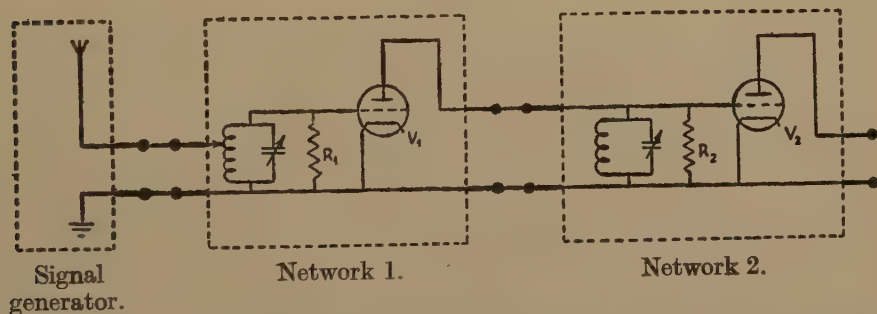


This result is of considerable value and may be extended to cover any number of networks in cascade ; thus :

$$F_{1+2+\dots+n} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \dots + \frac{F_n - 1}{G_1 G_2 \dots G_{n-1}} \quad \dots \quad (8)$$

This indicates immediately the well-known result that if the gain (G_1) of the first stage in a complex receiver is sufficiently great, then the overall noise figure is just the noise figure of the first stage (F_1).

Fig. 5.



Consider now the simple two-stage receiver in fig. 5. To simplify matters let the aerial, as before, be matched to the input circuit of V_1 and let the output impedance, \hat{r} , of V_1 be matched to the unloaded dynamic resistance, R_2 , of the coupling circuit (*i.e.* $\hat{r} = R_2$). Fig. 6 is then an equivalent circuit for this receiver, R_{eq1} accounts for the shot noise in V_1 and R_{eq2} for that arising in V_2 .

As derived previously :

$$F_1 = 2 + \frac{4}{R_1} \cdot R_{eq1} \quad (\text{Since } R_s = R_1.)$$

$$G_1 = \frac{\mu^2 R_1}{4R_2} \quad (\text{Since } R_g = R_1; \hat{r} = R_2.)$$

The derivation of F_2 follows precisely the same lines as that of F_1 , and hence by direct comparison :

$$F_2 = 2 + \frac{4}{R_2} \cdot R_{eq2}.$$

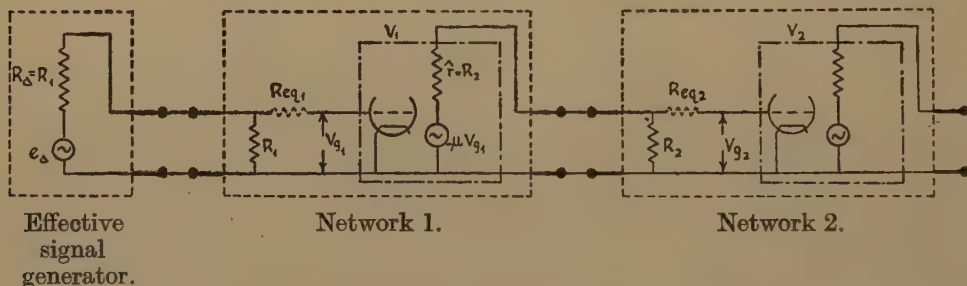
$$\begin{aligned} \text{Therefore : } F_{1+2} &= 2 + \frac{4}{R_1} \cdot R_{eq1} + \left[\frac{1 + (4R_{eq2}/R_2)}{\mu^2 R_1} \right] \cdot 4R_2 \\ &= 2 + \frac{4}{R_1} \cdot \left\{ R_{eq1} + \left[(R_2/4 + R_{eq2}) / \frac{\mu^2}{4} \right] \right\}. \quad \dots \quad (9) \end{aligned}$$

Formula (9) is then the transformation of the fundamental formula (7) into practical terms for this particular case.

4.2. "Secondary" Derivation of Noise Figure.

We shall consider now the derivation of a noise figure for the simple two-stage receiver in fig. 5, starting from the "engineering" definition.

Fig. 6.



If we refer back to the latter part of section 2.2 and equation (4) in particular, repeated here for convenience :

$$F = \frac{e_a^2}{e_g^2} \cdot \frac{(R_{eq} + R_g)}{R_a}, \quad \dots \quad (4)$$

we find that the general argument is applicable in all respects to the case in hand. The problem resolves itself into the determination of the effective R_{eq} to be considered as at the grid of V_1 to account for all the noise, referred to that point, apart from that generated by the effective thermal resistance R_g in the grid circuit of V_1 .

The shot noise in V_1 is accounted for by R_{eq1} "at" the grid of V_1 , so this must certainly be one component of R_{eq} ; i.e. denoting this component of R_{eq} by $R(s1)$ we have $R(s1) = R_{eq1}$.

Shot noise in V_2 is accounted for by R_{eq2} at the grid of V_2 , and it is clear that the component of R_{eq} at the grid of V_1 to account for, or replace, this will be R_{eq2} divided by the voltage gain squared from grid to grid (since an equivalent noise resistance is proportional to the effective mean square noise voltage at that point). Again, calling this

component $R(s2)$ we have $R(s2)=R_{eq2}/\dot{g}^2$, where \dot{g} =grid-to-grid voltage gain.

If now the effective resistance to a signal applied between grid and cathode of V_2 is R_{g2} (which in this case is $R_2/2$ since $\hat{r}=R_2$), it is tempting to say that the component of R_{eq} due to the thermal noise in the input of V_2 (say $R(t2)$) is similarly $R(t2)=R_{g2}/\dot{g}^2$.

Let us now see what result this would lead to. The grid-to-grid voltage gain is easily shown for the case chosen to be :

$$\dot{g}=\mu/2.$$

Hence :

$$\begin{aligned} R_{eq} &= R(s1) + R(t2) + R(s2). \\ &= R_{eq1} + \frac{R_{g2} + R_{eq2}}{\dot{g}^2} \\ &= R_{eq1} + [(R_2/2 + R_{eq2})/(\mu^2/4)]. \quad (\text{Since } R_{g2}=R_2/2.) \end{aligned}$$

Therefore from (4) :

$$\begin{aligned} F_{1+2} &= 4 \cdot \left\{ \frac{[R_{eq1} + (R_2/2 + R_{eq2})/(\mu^2/4)] + R_1/2}{R_1} \right\} \quad \left(\begin{array}{l} \text{Since : } R_g = R_1/2 \\ R_a = R_1 \\ e_s = e_a/2. \end{array} \right) \\ &= 2 + \frac{4}{R_1} \left\{ R_{eq1} + \left[(R_2/2 + R_{eq2}) \left/ \frac{\mu^2}{4} \right] \right\}. \quad \dots \dots (10) \end{aligned}$$

Comparison of (10) and (9) shows equivalence apart from the discrepancy in the denominator of the R_2 term. This indicates faulty reasoning in the computation of $(R(t2))$ above, since we may feel confident that the derivation from the fundamental formula is not at fault.

5. Examination of the Error.

It would appear that this error has been committed more than once in the past ; for example, E. W. Herold, in an otherwise excellent article⁽³⁾ published in 1942, deals with this particular question in the section entitled "Effect of output circuit and Second Tube on total equivalent noise resistance, R_{eq} ."

Quoting from this section :

"With a single tuned (coupling) circuit of resonant impedance \dot{R}_2 * and with a second tube of equivalent noise resistance R_{eq2} the mean squared noise at the grid of the second tube is proportional to $(\dot{R}_2 + R_{eq2})$. This may be referred to the grid of the first tube by dividing by the square of the grid-to-grid gain. The total equivalent noise resistance at the grid of the first tube is then :

$$R_{eq} = R_{eq1} + \frac{\dot{R}_2 + R_{eq2}}{\dot{g}^2},$$

where R_{eq1} is the equivalent noise resistance of the first tube."

* Herold uses the symbol R_2 in his article, but we have used \dot{R}_2 to avoid confusion with R_2 used in this note to signify the unloaded dynamic impedance of the coupling circuit.

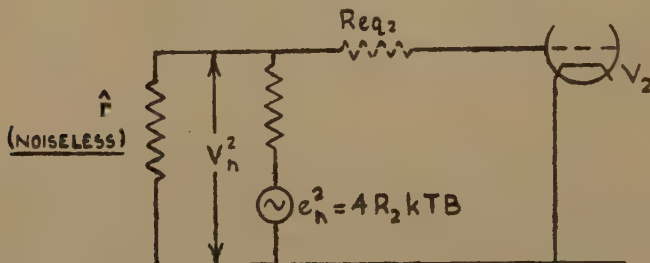
Herold clearly intends R_2 to include the shunting effect of the slope resistance of V_1 (\hat{r}) as modifying the unloaded dynamic impedance of the coupling circuit; *i. e.* $\hat{R}_2 = R_{g2}$ in our notation. Thus, in the case we have considered, $\hat{R}_2 = R_2/2$ since $R_2 = \hat{r}$. Hence we arrive at the same expression for R_{eq} as derived in section 4.2, namely:

$$R_{eq} = R_{eq1} + [(R_2/2 + R_{eq2})/(\mu^2/4)] \quad (\text{Since } \dot{g} = \mu/2.)$$

which we have already seen to be in error.

The source of error lies in the assumption that the resistance R_{g2} (or \hat{R}_2) is a source of thermal noise; part of it, R_2 —the unloaded dynamic impedance—certainly is, but \hat{r} is *not* a thermal noise source. This is immediately evident when we recollect that R_{eq1} already accounts for the noise generated in the valve, V_1 , and thus \hat{r} must be regarded as a *noiseless* shunting resistance. Fig. 7 shows the true state of affairs

Fig. 7.



as regards thermal noise generation, in the input circuit of V_2 . We see immediately that in this case:—

$$v_n^2 = \frac{e_n^2 \hat{r}^2}{(\hat{r} + R_2)^2},$$

and if

$$\hat{r} = R_2$$

$$v_n^2 = \frac{e_n^2}{4} = 4 \cdot \frac{R_2}{4} \cdot kTB \quad \dots \dots \dots (11)$$

as compared with

$$v_n^2 = 4 \cdot \frac{R_2}{2} \cdot kTB, \quad \dots \dots \dots (12)$$

if \hat{r} were thermally “noisy.”

Clearly then, from (11) the true value of effective thermal resistance at the input of V_2 in this case is $\frac{R_2}{4}$ and not $\frac{R_2}{2}$ as indicated by the faulty equation (12). This value of R_2 then agrees with that derived from the fundamental formula as in equation (9). This agreement is, of course, perfectly general.

It is hoped that this discussion may be of some value both in pointing out this particular error and in demonstrating the use and value of the fundamental formula.

References.

- (1) "Available" Signal power :—
This is the maximum available external power obtainable from a signal source, *i. e.* the power delivered to a matched load whose complex impedance is equal to the conjugate of the source output impedance.
- (2) "Available" Noise power :—
This is the maximum available external noise power obtainable from a noise source, *i. e.* the noise power delivered to a matched *noiseless* load within a specified band-width.
- (3) E. W. Herold, "An Analysis of the Signal-to-Noise Ratio of U.H.F. Receivers." R.C.A. Review, pp. 302-331 (January, 1942).

XLVI. *The Accumulation of Chance Effects and the Gaussian Frequency Distribution.*

By LUDWIK SILBERSTEIN, Ph.D.*

[Received February 18, 1944.]

IT has often been surmised, either quantitatively or in a more or less vague qualitative way, that Gauss's law of error or the so-called normal frequency distribution of deviations from a mean would be produced by the co-operation of a great number of independent causes, if the effect of each of these taken separately is uncertain within some limits. In the research field of photography, this question is of particular interest in connection with the size-frequency distribution among the grains or silver halide crystals of an emulsion. A clear insight into the *modus operandi* of such a plurality of causes may also be useful in a variety of other situations.

The object of the present paper is to determine rigorously the probability of any resultant effect of a number n of independent causes or processes, as the accumulation of their individual chance effects, and to investigate the limiting form to which this probability function tends with increasing n .

Let the effect of a process, in a single trial, be equally likely to have *any value* between 0 and 1, so that the probability of an effect x to $x+dx$ is just dx and the probable effect of any single process is $\int_0^1 x dx$ or $\frac{1}{2}$. What is the probability, $p_n(x)dx$, that the resultant effect of an n -fold repetition of the process, or a succession of n trials, shall fall between x and $x+dx$?

By the definition of our elementary process,

$$p_1(x)=1 \text{ for } 0 < x < 1 \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and nil elsewhere, and the probability functions or "frequencies" $p_2(x)$, $p_3(x)$, and so on have to be derived successively from this datum.

* Communication No. 957 from the Kodak Research Laboratories.

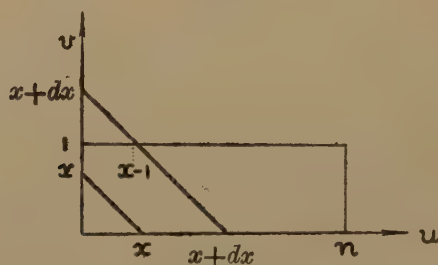
For this purpose we require a recurrency relation between $p_{n+1}(x)$ and $p_n(x)$.

Let $u+du$ be the resultant effect of n trials and $v+dv$ the effect of one more (single) trial. The sum of these effects has to fall between x and $x+dx$. Thus

$$p_{n+1}(x) dx = \iint p_n(u) du dv,$$

to be integrated over a domain satisfying the last requirement. If we think of u, v as the Cartesian co-ordinates of a point, then this domain is a strip between the straight lines $u+v=x$ and $u+v=x+dx$, contained within the rectangle of base n and height 1 (see fig. 1). The strip being equally inclined to the axes, its width is $dx/\sqrt{2}$ and its length element $\sqrt{2} du$; hence, its element of area is $du dx$. Thus, $p_{n+1}(x) = \int p_n(u) du$. The upper limit of u is x (or n , if x exceeds n) and its lower limit is $x-1$ (or 0, if $x < 1$). Since $p_n(u)$ vanishes beyond $u=n$ and below $u=0$,

Fig. 1.



we may always extend the integration from $x-1$ to x (regardless of the bracketed clauses), for the added intervals contribute nothing to the integral.

Thus, the fundamental recurrency relation becomes

$$p_{n+1}(x) = \int_{x-1}^x p_n(u) du. \quad \dots \quad (1)$$

It holds for any number of processes, from $n=1$ onwards. This relation, together with $p_1(x)$ as given by (1), will determine uniquely the whole set of probability functions, namely, by successive quadratures of polynomials.

Before passing to generalities, let us determine these functions for the first few values of the integer n .

Let $n=2$.—The relation (I) becomes

$$p_2(x) = \int_{x-1}^x p_1(u) du,$$

where $p_1=1$ for $0 < u < 1$ and nil elsewhere. Thus, if $0 < x < 1$, we have

$$p_2(x) = \int_0^x du = x,$$

and if $1 < x < 2$, then

$$p_2(x) = \int_{x-1}^1 du = 2-x.$$

(If $x > 2$, then also $x-1 > 1$ and $p_2=0$, of course.) In fine, the whole probability function consists of two branches, stretching from $x=0$ to 1 and from 1 to 2:

$$p_2(x) = x, \quad 2-x. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

It is continuous at $x=1=\frac{n}{2}$, where it attains its maximum, $P_2=1$. The corresponding "curve" consists of two straight lines, *symmetrical* with respect to the peak (a sharp point) at $x=1$. This, the most probable value of x , is also its *probable* value $[x]$. For we have

$$[x] = \int_0^2 x p_2(x) dx = \int_0^1 x^2 dx + \int_1^2 x(2-x) dx = \frac{1}{3} + 3 - \frac{7}{3} = 1,$$

as could be expected. We shall see presently that, generally, $[x] = \frac{n}{2}$. With equal ease the reader will find $[x^2] = \frac{7}{6}$, so that $[x^2] - [x]^2 = \frac{1}{6}$, and the probable value of x with its $P.E_1$ is

$$x = 1 \pm \frac{1}{\sqrt{6}}.$$

We may now pass on to the next higher case.

Let $n=3$.—The function

$$p_3(x) = \int_{x-1}^x p_2(u) du$$

will consist of three branches extending from $x=0$ to 1, from 1 to 2, and from 2 to 3. The first branch is, by (2),

$$p_3(x) = \int_0^x u du = \frac{1}{2}x^2.$$

The second branch is, again by (2) and after simple reductions,

$$p_3(x) = \int_{x-1}^1 u du + \int_1^x (2-u) du = -x^2 + 3x - \frac{3}{2} = \frac{3}{4} - \left(x - \frac{3}{2}\right)^2,$$

and the third branch (since p_2 vanishes beyond $x=2$),

$$p_3(x) = \int_{x-1}^2 (2-u) du = \frac{1}{2}x^2 - 3x + \frac{9}{2} = \frac{1}{2}(3-x)^2.$$

Thus the complete function consists of the three branches

$$p_3(x) = \frac{1}{2}x^2, \quad \frac{3}{4} - \left(x - \frac{3}{2}\right)^2, \quad \frac{1}{2}(3-x)^2, \quad . \quad . \quad . \quad (3)$$

passing into each other continuously at $x=1$ and $x=2$. The corresponding curve (shown in fig. 2) consists of three parabolic arcs. It is again symmetrical with respect to its peak, $P_3 = \frac{3}{4}$, placed at $x = \frac{3}{2} = \frac{n}{2}$.

In other words, $p_3(x)$ is an even function of $x - \frac{3}{2}$. The probable value of x is again readily verified to be $[x] = \frac{3}{2}$.

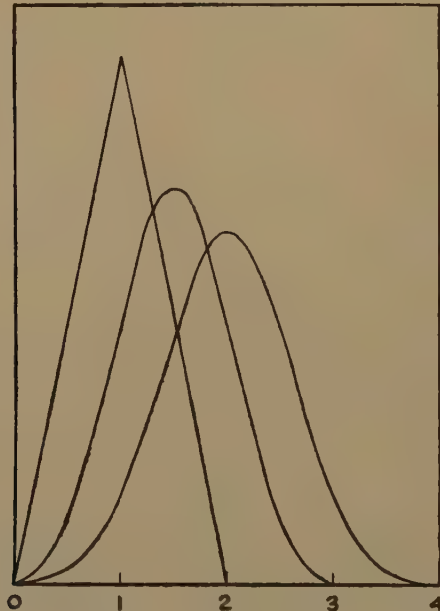
Let $n=4$. The four branches of the probability function are derived from (3) in much the same way as before. Everything is again symmetrical with respect to $x = \frac{n}{2} = 2$. It will be convenient to introduce the variable

$$\xi = x - \frac{n}{2}.$$

If $|\xi|$ stands for its absolute value, then the first two branches of the function are

$$p_4(x) = \frac{1}{3!} (2 - |\xi|)^3, \quad \frac{2}{3} - \xi^2 + \frac{1}{2} |\xi|^3, \quad . \quad . \quad . \quad . \quad (4)$$

Fig. 2.



the third branch is identical with the second, and the fourth branch with the first. The curve (fig. 2) is symmetrical with respect to its peak, $P_4 = \frac{2}{3}$, placed at $\xi = 0$. As before, $[x] = \frac{n}{2} = 2$. In terms of ξ , the preceding two probability functions may be rewritten thus :

$$p_2(x) = 1 - |\xi|, \quad 1 - |\xi|, \quad . \quad . \quad . \quad . \quad (2')$$

$$p_3(x) = \frac{1}{2!} \left(\frac{3}{2} - |\xi| \right)^2, \quad \frac{3}{4} - \xi^2, \quad \frac{1}{2!} \left(\frac{3}{2} - |\xi| \right)^2, \quad . \quad . \quad . \quad (3')$$

We may quote the results for just one more case.

Let $n=5$. The first three branches of the function are

$$p_5(x) = \frac{x^4}{4!}, \quad \frac{55}{96} + \frac{5}{24}|\xi| - \frac{5}{4}\xi^2 + \frac{5}{6}|\xi|^3 - \frac{1}{6}\xi^4, \quad \frac{115}{192} - \frac{5}{8}\xi^2 + \frac{1}{4}\xi^4, \quad (5)$$

the fourth branch coincides with the second, and the fifth is $\frac{1}{4!}(5-x)^4$.

The corresponding curve, much of the same character as the preceding one, is symmetrical about its peak $P_5 = \frac{115}{192}$, placed at $\xi=0$ or $x=[x] = \frac{5}{2}$.

It is manifest from the mechanism of derivation that the first branch is, for any n ,

$$p_n(x) = \frac{x^{n-1}}{(n-1)!} \quad (0 < x < 1) \quad . \quad . \quad . \quad (a)$$

and the last branch,

$$p_n(x) = \frac{(n-x)^{n-1}}{(n-1)!} \quad (n-1 < x < n) \quad . \quad . \quad . \quad (b)$$

The remaining branches are polynomials of degree $n-1$ in x or ξ , with numerical coefficients varying in an intricate manner from case to case. The peak of the probability curve is gradually depressed with increasing n ; in fact, as we saw,

$$P_2=1, \quad P_3=\frac{3}{4}, \quad P_4=\frac{2}{3}, \quad P_5=\frac{115}{192}, \quad . \quad . \quad . \quad (c)$$

to which I may add that $P_6 = \frac{83}{160}$ or a little above $\frac{1}{2}$, and P_7 is a little below $\frac{1}{2}$. Since the area under the whole curve or $\int_0^n p_n(x) dx$ is, by construction, always equal to 1, a depression of its ordinates with increasing n could be expected beforehand.

That the symmetry about $x = \frac{n}{2}$, noticed in all the preceding cases, holds for any value of n , may be proved rigorously as follows :

Suppose that it holds for a given n , i. e., that

$$p_n\left(\frac{n}{2} + \alpha\right) = p_n\left(\frac{n}{2} - \alpha\right)$$

for any α . Then, by the fundamental relation (I),

$$p_{n+1}\left(\frac{n+1}{2} + \alpha\right) = \int_{\frac{n}{2} + \alpha - \frac{1}{2}}^{\frac{n}{2} + \alpha + \frac{1}{2}} p_n(u) du$$

and

$$p_{n+1}\left(\frac{n+1}{2} - \alpha\right) = \int_{\frac{n}{2} - \alpha - \frac{1}{2}}^{\frac{n}{2} - \alpha + \frac{1}{2}} p_n(u) du.$$

The two integration intervals are equal (viz, each equal to 1) and sym-

metrically placed with respect to $u = \frac{n}{2}$; hence the two integrals are equal and

$$p_{n+1}\left(\frac{n+1}{2} + \alpha\right) = p_{n+1}\left(\frac{n+1}{2} - \alpha\right),$$

that is to say, if the symmetry holds for n , it holds also for $n+1$. But it certainly holds for $n=2$. It must hold, therefore, for any number n of processes.

In other words, any $p_n(x)$ is an *even* function of $\xi = x - \frac{n}{2}$.*

The probable resultant of n processes is defined by

$$[x] = \int_0^n x p_n(x) dx = \int_{-n/2}^{n/2} \left(\frac{n}{2} + \xi\right) p_n(\xi) d\xi,$$

and since $p_n(\xi) = p_n(-\xi)$, this becomes at once

$$[x] = n \int_0^{n/2} p_n(\xi) d\xi = \frac{1}{2}n,$$

as stated before.

The formulæ (2) to (5) and the corresponding graphs may suffice for forming an idea about the shape of the probability curves in the case of the first few values of the number n . Each of these formulæ consists of n distinct algebraic expressions varying in an intricate way from branch to branch and from one curve to the next higher one. They are thus, manifestly, unsuitable for disclosing the general structure of the probability function $p_n(x)$ in its dependence on the number n and, especially, its properties for great values of this number.

In order to find for this function a single analytical expression covering for a given n , all of its branches and valid for all n , the form of our starting datum, $p_1(x)$, has to be suitably modified. The three branches of $p_1(x)$ actually contained in formula (1) and its clause (nil for all $x < 0$, unity for $0 < x < 1$, nil for all $x > 1$) have to be replaced by a unique expression valid for all real x , from $-\infty$ to ∞ . Such an expression is the corresponding Fourier integral. Since p_1 is an even function of $\xi = x - \frac{1}{2}$, Fourier's theorem gives

$$p_1(\xi) = \frac{2}{\pi} \int_0^\infty \cos v\xi \cdot dv \int_0^\infty p_1(u) \cos vu \cdot du.$$

The second integral is, simply,

$$\int_0^1 \cos vu \cdot du = \frac{1}{v} \sin \frac{v}{2},$$

so that

$$p_1(\xi) = \frac{2}{\pi} \int_0^\infty \frac{\sin \frac{v}{2}}{v} \cos(\xi v) dv,$$

* This even function of ξ depends, besides, in some intricate way on n itself. The form of this dependence will be discussed hereafter.

or, replacing throughout $v/2$ by v ,

$$p_1(x) = \frac{2}{\pi} \int_0^\infty \frac{\sin v}{v} \cos 2\left(x - \frac{1}{2}\right)v \cdot dv. \quad (1')$$

This is the required expression of our starting function, representing it correctly for all x whatever*.

Let us now apply to the function $p_1(x)$ thus expressed the recurrency relation (I). The order of integrations over v and over x is interchangeable. Keeping this in mind, let us write, for the moment,

$$p_n(x) = \frac{2}{\pi} \int_0^\infty \frac{\sin v}{v} f_n dv, \quad (6)$$

so that

$$f_1(x) = \cos 2v\left(x - \frac{1}{2}\right),$$

and, by (I),

$$f_{n+1}(x) = \int_{x-1}^x f_n(u) du. \quad (7)$$

Thus,

$$f_2(x) = \int_{x-1}^x \cos 2v\left(u - \frac{1}{2}\right) du = \frac{1}{2v} \left\{ \sin 2v\left(x - \frac{1}{2}\right) - \sin 2v\left(x - \frac{3}{2}\right) \right\},$$

or, after simple trigonometrical transformations,

$$f_2(x) = \frac{\sin v}{v} \cos 2v(x-1).$$

We see that, in passing from f_1 to f_2 , $x - \frac{1}{2}$ is replaced by $x - \frac{2}{2}$ and the factor $\frac{\sin v}{v}$ is introduced. This suggests, for any n ,

$$f_n(x) = \left(\frac{\sin v}{v}\right)^{n-1} \cos 2v\left(x - \frac{n}{2}\right). \quad (8)$$

To prove this property, it is enough to show that, if it holds for n , it will hold also for $n+1$. Now by (7), we have

$$f_{n+1}(x) = \left(\frac{\sin v}{v}\right)^{n-1} \int_{x-1}^x \cos 2v\left(u - \frac{n}{2}\right) du.$$

The integral is readily found to be

$$\frac{1}{2v} \left\{ \sin 2v\left(x - \frac{n}{2}\right) - \sin 2v\left(x - \frac{n}{2} - 1\right) \right\} = \frac{\sin v}{v} \cos 2v\left(x - \frac{n+1}{2}\right),$$

so that

$$f_{n+1}(x) = \left(\frac{\sin v}{v}\right)^n \cos 2v\left(x - \frac{n+1}{2}\right),$$

as in (8), with n replaced by $n+1$. And since (8) holds for $n=1$, it will hold for any n . It remains to substitute (8) in (6).

Ultimately, therefore, the probability function or frequency is rigorously expressed by

$$p_n(x) = \frac{2}{\pi} \int_0^\infty \left(\frac{\sin v}{v}\right)^n \cos 2\left(x - \frac{n}{2}\right)v \cdot dv. \quad (II)$$

* At a point of discontinuity, the Fourier integral is well known to give the arithmetical mean of the actual values of the function, on both sides of the point, that is, in our case, $p_1(0) = \frac{1}{2}(0+1) = \frac{1}{2}$ and similarly $p_1(1) = \frac{1}{2}$. This, however, is immaterial for the subsequent integrations over x .

This covers at once all branches, in fine, the complete probability curve, and holds for any number of processes, from $n=1$ onwards. The reader may easily verify the fact that our previous algebraic expressions (2), (3), etc., when translated into Fourier integrals, are all identical with (II) for $n=2, 3$, and so on. We now see, once more, that the frequency is an even function of $\xi = x - \frac{n}{2}$, the deviation from the mean or the probable effect $\frac{n}{2}$,

$$p_n(\xi) = \frac{2}{\pi} \int_0^\infty \left(\frac{\sin v}{v} \right)^n \cos(2\xi v) dv.$$

It depends, besides, on the number of processes through the exponent of the first factor of the integrand. The maximum of any frequency curve, reached at $\xi=0$, is expressed by the integral

$$P_n = \frac{2}{\pi} \int_0^\infty \left(\frac{\sin v}{v} \right)^n dv. \quad \dots \dots \dots (9)$$

Its value for $n=1$ is well known to be $P_1=1$, in agreement with the original definition of $p_1(x)$. From our algebraic solutions we may conclude that such also is its value for $n=2$, and that $P_3=\frac{3}{4}$, $P_4=\frac{2}{3}$, $P_5=\frac{115}{192}$. So far as we know, these integrals have not been previously determined. For higher values of n , the integral could be evaluated by mechanical quadrature, but the necessity of doing so will be obviated presently.

The rigorous frequency formula (II) which, theoretically, is the complete solution of the problem, cannot be evaluated in finite form in terms of ξ and n without falling back on the successive algebraic integrations applied at the outset. It has, however, the advantage of being perfectly general and thus suitable for showing us the properties of the frequency at high values of n which are of especial interest in any application. Clearly, no such result could be achieved by extrapolation from the algebraic solutions for the first few values of n found before.

If n is a great number (such as 1000 or even 100), the only relevant contributions to the integral in (II) are those due to the closest neighbourhood of $v=0$, where $\frac{\sin v}{v}$ is not much smaller than unity. Now, in this neighbourhood, we may write, up to v^4 -terms,

$$\frac{\sin v}{v} = 1 - \frac{v^2}{6} \doteq e^{-v^2/6}, \quad \left(\frac{\sin v}{v} \right)^n = e^{-\frac{nv^2}{6}}.$$

Accordingly, the upper limit of the integral should be replaced by some low or moderate value $v=\alpha$. But, since the contributions beyond α would be insignificant (the more so, the greater n), we may as well retain the original upper limit of the integral, a somewhat arbitrary procedure to be tested *a posteriori*. In this manner the solution (II) becomes, for great n ,

$$p_n(\xi) = \frac{2}{\pi} \int_0^\infty e^{-\frac{nv^2}{6}} \cos(2\xi v) dv,$$

The second factor has been left untouched. The approximate legitimacy of the mutilation of the first factor of the integrand may be tested by noticing that the maximum frequency, rigorously expressed by (9), is now cut down to

$$P_n = \frac{2}{\pi} \int_0^\infty e^{-\frac{nv^2}{6}} dv = \sqrt{\frac{6}{\pi n}}.$$

Curiously enough, this formula is, even for low values of n , not far off the mark. In fact, apart from P_1 , which it overshoots by 0.38, it gives $P_2=0.978$ against the rigorous unity, $P_3=0.798$ against 0.750, $P_4=0.691$ against 0.667, and $P_5=0.618$ against $\frac{115}{192}$ or 0.599. Such being the case, we may accept the last expression for $p_n(\xi)$ with confidence.

The integral is instantly evaluated*, and the approximate expression for the frequency, applicable to great and, as we saw, even to moderate numbers n , becomes

$$p_n(\xi) = \sqrt{\frac{6}{\pi n}} e^{-\frac{6}{n}\xi^2}, \quad \dots \dots \dots (10)$$

where $\xi = x - \frac{n}{2} = x - \bar{x}$. It is precisely of the form of Gauss's law of error.

The frequency curve which, for $n=2$, consisted of a pair of straight lines with a sharp point as peak, is, with increasing n , smoothed out to the familiar "normal" curve, symmetrical about the mean resultant effect, $\bar{x}=n/2$. Rigorously, of course, the probability function or the frequency vanishes outside the interval $\xi = \pm \frac{n}{2}$ and we have, strictly, $\int_{-n/2}^{n/2} p d\xi = 1$, while the curve (10) stretches indefinitely on both sides and its area above that interval is short of unity. The defect, however, is quite insignificant even for moderate values of n . In fact, by (10), this area

amounts to $\Phi\left(\sqrt{\frac{3\pi}{2}}\right)$, where Φ is the error-function. Hence, even for $n=6$ only, it is as close to unity as 0.99998.

The standard deviation σ is defined by $\sigma^2 = \int \xi^2 p_n(\xi) d\xi$. Thus, for the frequency distribution (10),

$$\sigma^2 = \frac{n}{3\sqrt{\pi}} \int_0^{\sqrt{\frac{3n}{2}}} u^2 e^{-u^2} du.$$

If n exceeds but a few units, the integral amounts to $\frac{\sqrt{\pi}}{4}$, so that the standard deviation is

$$\sigma = \sqrt{\frac{n}{12}}, \quad \dots \dots \dots (11)$$

* As the real part of $\int_0^\infty e^{2i\xi v - nv^2/6} dv$.

and directly from the mass-absorption coefficients of the constituents and the geometry of the system.

The micro-absorption factor has been consistently omitted in X-ray investigations, and it is highly probable that phenomena which have been attributed to other causes may, in reality, be ascribed to differential absorption effects. For example, measurements of the X-ray intensities reflected into the Debye-Scherrer spectra of filed metal powders reveal apparently diminished values of the atomic scattering factors. This has been attributed to random displacements of the atoms about their mean lattice positions. We shall indicate in Section I. how the dimensions of the reflecting crystallites can lead to intensity changes of the same order of magnitude, so that effects ascribed to lattice distortion may, in reality, be inherent in the methods of standardizing the X-ray reflections.

In Section II. it will be shown how the micro-absorption factor may be used as a criterion for the choice between two theories concerning the broadening of the superlattice lines in Cu_3Au .

I. *Atomic Scattering Factors from Cold-worked Copper and Nickel.*

G. W. Brindley and F. W. Spiers⁽³⁾ have compared the atomic scattering factors f of copper and nickel powders obtained by using unannealed filings with the scattering factors derived from the chemically precipitated metals. They found that the filings gave f values which were several per cent. lower than those of the precipitates. The curves in fig. 1 reproduced from their paper show the extent of the deviations.

The absolute f -values for the precipitated powders were obtained by standardizing the intensities of the reflections against those of finely divided KCl. Using the standardized intensities from the precipitated powders, the intensities from the filings were then put upon an absolute basis. The lower f values for the filings calculated from the standardized intensity measurements were attributed to random atomic displacements which had the same general effect on the scattering factors as the thermal vibrations of the lattice. The lower values could therefore be accounted for by an expression of the form

$$f'_T = f_T e^{-M}, \quad \dots \quad (1)$$

where

$$M = \frac{8\pi^2}{3} \left(\frac{\sin \theta}{\lambda} \right)^2 \overline{u^2},$$

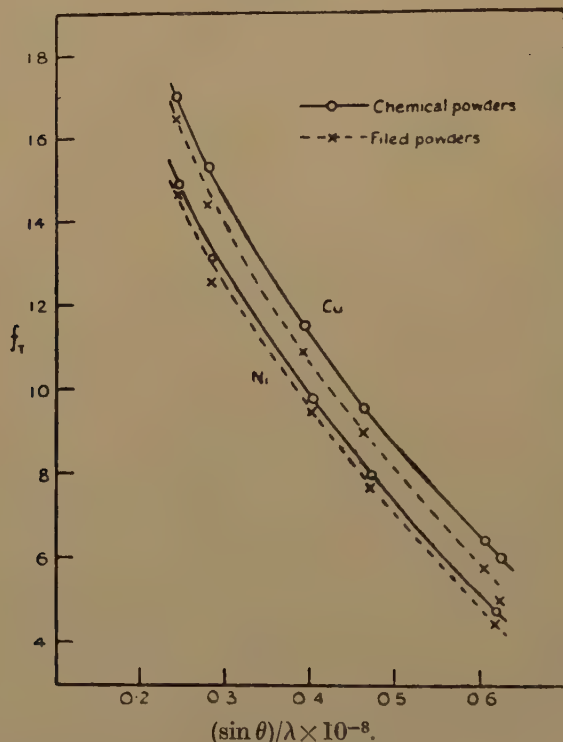
$\overline{u^2}$ being the mean square of the average atomic displacement, f_T the atomic scattering factor for the undistorted lattice and f'_T that for the distorted lattice.

In their work, the powders were examined under the microscope. For chemically precipitated copper, the size of the powder particles was about 5×10^{-4} cm., while for filed chill-cast copper the size was 2×10^{-3} cm. On the basis of these measurements, it was assumed by Brindley and Spiers that the grain size was too small to have any appreciable effect

on the intensities of the X-ray reflections. Now the microscope does not, in such cases, yield the true dimensions of the reflecting crystallites, but only their aggregate size. The true size of the individual crystallites in the chemical precipitate probably did not exceed a value of 10^{-5} cm., while the crystallite size in the filings was more probably of the order of 10^{-4} cm., a value which is still larger than 0.7×10^{-5} cm. given by W. A. Wood as the lower limiting size for crystals in severely cold-worked copper of high purity.

In the absence of any X-ray measurements to determine the true crystallite size, the actual dimensions must remain a matter of speculation.

Fig. 1.



Atomic scattering factors for chemically prepared and unannealed filed powders of copper and nickel at room temperature T (G. W. Brindley and F. W. Spiers, *Phil. Mag.* xx. p. 882 (1935)).

Nevertheless, it can be stated that only when the dimensions of the crystallites fall below 10^{-5} cm. can the effects of X-ray absorption be neglected.

For a flat stationary layer of powder of effectively infinite thickness, the total energy I reflected into the Debye-Scherrer halo is given by the equation

$$I = I_0 \frac{\lambda^3 e^4}{2m^2 c^4} \cdot \rho \cdot \frac{N^2 F^2}{\mu'} \left(\frac{1 + \cos^2 2\theta}{\sin \theta \sin 2\theta} \right) \frac{\rho'}{\rho} \left[\frac{\sin (2\theta - \psi)}{\sin (2\theta - \psi) + \sin \psi} \right] \tau, \quad (2)$$

which may be abbreviated to

$$I = \text{constant} \cdot p \cdot \frac{N^2 F^2}{\mu'} \frac{\rho'}{\rho} \phi(\theta) G \tau, \quad (3)$$

where p = number of cooperating planes of form hkl ,

N = number of unit cells per unit volume of crystal,

F = the structure factor,

θ = the Bragg angle,

ψ = angle of setting of powder plate with respect to incident X-ray beam,

μ' = effective linear absorption coefficient of powder block,

ρ' = effective density of powder block,

ρ = true density of the substance itself,

τ = micro-absorption factor for the reflecting crystallite.

The trigonometrical functions are abbreviated to the term $\phi(\theta)$ and G is the geometrical factor, taking into account the fact that different reflections leave the powder block at different angles.

If we compare a given reflection from the filed powder with the corresponding reflection from the chemically prepared powder, p , N , $\phi(\theta)$, G and μ' are the same. Also $F = 4f_T$ for face-centred cubic lattices, so that the ratio of the intensities of a given reflection from the two powders becomes

$$\frac{I_1}{I_2} = \left(\frac{f_{T1}}{f_{T2}} \right)^2 \frac{\tau_1}{\tau_2} \quad (4)$$

Brindley and Spiers neglect the micro-absorption factors and assume the ratio to take the form

$$\frac{I_1}{I_2} = \left(\frac{f_{T1}}{f_{T2}} \right)^2 \quad (5)$$

In the absence of any real data on the true crystallite sizes, we shall use the arbitrary values $a = 10^{-5}$ cm. and $a = 2 \times 10^{-4}$ cm. for the radii of the precipitated crystallites and filings from chill-cast copper respectively. To obtain the τ values, we first of all compute the values μa for the particles and then look up the tables. With $\text{CuK}\alpha$ radiation, $\frac{\mu}{\rho} = 52.7$, and since $\rho = 8.9$ we have $\mu = 469$. For the precipitate $\mu a = 0.00469$ and for the filings $\mu a = 0.0938$. Looking up the tables, 100τ (precipitate) ≈ 100 , and 100τ (filings) ≈ 88 . For $\mu a < 0.1$, the variation in 100τ with the Bragg angle, θ , is less than 0.7 per cent., and so we may neglect it in the foregoing calculations.

Equation (4) yields

$$\begin{aligned} f_T(\text{filings}) &= \sqrt{\frac{I(\text{filings})}{I(\text{pptd.})}} \cdot f_T^*(\text{pptd.}) \sqrt{\frac{100}{88}} \\ &= 1.07 f_T^*(\text{filings}), \end{aligned}$$

where f_T^* (filings) is the value obtained by Brindley and Spiers using

equation (5). By making an allowance for the τ factor, we can *raise* the value of f_T^* (filings) by 7 per cent. and obtain a close agreement with the f_T (pptd.) values, instead of assuming an arbitrary e^{-M} factor which *reduces* f_T (pptd.) to the lower f_T^* (filings) values. A comparison of the two types of correction for chill-cast copper is made in Table I. The first three sections in the table are Brindley and Spiers' experimental results. The last section contains the author's adjustments for micro-absorption.

TABLE I.

Comparison of f_T values for filed chill-cast and chemically precipitated copper.

$$\begin{aligned} a \text{ (filings)} &= 2 \cdot 10^{-4} \text{ cm.} & a \text{ (pptd.)} &= 10^{-5} \text{ cm.} \\ 100\tau \text{ (filings)} &= 88. & 100\tau \text{ (pptd.)} &= 100. \end{aligned}$$

$$\sqrt{\frac{100}{88}} = 1.07.$$

Reflec- tion.	$\frac{\sin \theta}{\lambda}$	Not corrected.		Corrected.		Corrected.	
		f_T (filed).	f_T (pptd.).	f_T (filed).	$f_T e^{-M}$ (pptd.).	$f_T \sqrt{\frac{100}{88}}$ (filed).	f_T (pptd.).
111	0.241	16.5	17.0	16.5	16.95	17.6	17.0
200	0.278	14.4	15.3	14.4	15.05	15.4	15.3
220	0.393	10.9	11.5	10.9	11.0	11.55	11.5
311	0.461	8.95	9.5	8.95	8.9	9.6	9.5
331	0.605	5.8	6.4	5.8	5.75	6.2	6.4
420	0.621	5.0	6.0	5.0	5.35	5.35	6.0

It will be seen from Table I. that the agreements obtained by the two methods of correction are equally good.

It was found by Brindley and Spiers that the results for filings of hard-drawn copper wire, though still below those for precipitated copper, were, nevertheless, *higher* than those obtained from the filings of chill-cast copper. If cold-working is the cause of the reduced f values, we should have expected hard-drawn copper to give the *lowest* values. On the other hand, if we make the justifiable assumption that the crystallite size is *smaller* for hard-drawn copper than for the filings from the chill-cast variety, the τ factor becomes larger and the f values appear larger. This is borne out by the results in Table II. Results for filed and precipitated nickel are given in Table III. In all cases the agreements are as good when τ -factor corrections are made to the filings as when e^{-M} factors are applied to the precipitated powders. It would seem that the evidence tends to favour the micro-absorption effect as the reason for lower f values in cold-worked metal filings.

TABLE II.

Comparison of f_T values for filed hard-drawn copper wire and chemically precipitated copper.

$$a \text{ (filings)} = 1.5 \times 10^{-4} \text{ cm.} \quad a \text{ (pptd.)} = 10^{-5} \text{ cm.}$$

$$100\tau \text{ (filings)} = 91.7. \quad 100\tau \text{ (pptd.)} = 100.$$

$$\sqrt{\frac{100}{91.7}} = 1.045.$$

Reflec- tion.	$\frac{\sin \theta}{\lambda}$.	Not corrected.		Corrected.		Corrected.	
		f_T hard- drawn.	f_T (pptd.).	f_T hard- drawn.	$f_T e^{-M}$ (pptd.).	$f_T \sqrt{\frac{100}{91.7}}$ hard- drawn.	f_T (pptd.).
111	0.241	16.85	17.0	16.85	16.95	17.6	17.0
200	0.278	14.5	15.3	14.5	15.05	15.2	15.3
220	0.393	11.3	11.5	11.3	11.0	11.8	11.5
311	0.461	9.15	9.5	9.15	8.9	9.55	9.5
331	0.605	5.6	6.4	5.6	5.75	5.86	6.4
420	0.621	5.2 ^a	6.0	5.2	5.35	5.45	6.0

TABLE III.

Comparison of f_T values for filed and chemically precipitated nickel.

$$a \text{ (filings)} = 10^{-4} \text{ cm.} \quad a \text{ (pptd.)} = 10^{-5} \text{ cm.}$$

$$\mu_{Ni} = 442.8 \text{ for CuK}\alpha \text{ radiation.}$$

$$100\tau \text{ (filings)} = 94. \quad 100\tau \text{ (pptd.)} = 100. \quad \sqrt{\frac{100}{94}} = 1.03.$$

Reflec- tion.	$\frac{\sin \theta}{\lambda}$.	Not corrected.		Corrected.		Corrected.	
		f_T (filed).	f_T (pptd.).	f_T (filed).	$f_T e^{-M}$ (pptd.).	$f_T \sqrt{\frac{100}{94}}$ (filed).	f_T (pptd.).
111	0.246	14.6	14.9	14.6	14.75	15.0	14.9
200	0.284	12.6	13.1	12.6	12.9	13.0	13.1
220	0.402	9.5	9.8	9.5	9.5	9.8	9.8
311	0.471	7.65	7.95	7.65	7.65	7.9	7.95
331	0.619	4.45	4.75	4.45	4.45	4.6	4.75

II. The Causes of Line-broadening by the Superlattice of Cu_3Au .

The atomic rearrangements with superlattice formation in the alloy Cu_3Au have been studied by F. W. Jones and C. Sykes ⁽⁴⁾. With a suitable heat treatment they show that three processes occur during the atomic interchanges within the lattice, namely :

(a) Formation of nuclei which have an equilibrium degree of order at the temperature T.

(b) Growth of degree of order S within the nuclei as T is reduced.

(c) Growth of nuclei until they meet. Since the ordered regions do not match at the surfaces of junction, the lattice as a whole appears to be broken up into a "foam structure." These ordered cells are referred to as "anti-phase domains," and the lattice is said to possess short-distance order.

When the domains are less than 10^{-4} cm. in cross-section, the superlattice lines in the Debye-Scherrer patterns become diffuse and the dimensions of the domains may be computed by the well-known Scherrer expression

$$\beta = \frac{1.0 \lambda}{\epsilon \cos \theta},$$

where β = the angular broadening of the line in radians,

λ = the wave-length of the radiation,

ϵ = the dimensions of the domains measured in a direction perpendicular to the reflecting planes, and

θ = the Bragg angle.

During the process of ordering and nuclei growth, the main lines in the Debye-Scherrer pattern remain unchanged in width and in intensity.

This picture of nuclei growth is one which is relatively easy to grasp, and it seems to offer a perfectly satisfactory explanation for the broad superlattice reflections. However, a totally different mechanism may be the underlying cause. In cobalt the lattice is "faulted" at irregular intervals, and this gives rise to a series of sharp and diffuse haloes. On this basis, O. S. Edwards and H. Lipson ⁽⁵⁾ have expressed the view that there is no obvious justification for treating the broadening of the superlattice lines in Cu_3Au as being due to the small size of the anti-phase domains.

The question then arises: If the two theories are equally capable of explaining the broadening of the superlattice reflections, is there a criterion for distinguishing between them? It is felt that an experiment can be performed which will distinguish between the two possibilities.

If we consider the formation of the Debye-Scherrer pattern, the absorption factor $A = \tau\alpha$ will depend on the τ value for the reflecting crystals and on the α factor for the specimen as a whole. When the main lines (m) of the pattern are formed, the τ_m factor will depend on the size of the reflecting crystallite, but when the superlattice lines (s) are formed, the τ_s factor must now depend on the size of the reflecting domains. Thus the ratio τ_s/τ_m must vary with the size of the domains, the size of the reflecting crystals, and with the wave-length of the radiation which governs the value of μa from which τ is obtained.

On the other hand, if the line-broadening is attributable to "faulting" of the lattice, the variations in line width must depend entirely on the lattice as a whole. Thus the τ_s and τ_m factors should to be the same no matter what the breadth of the superlattice lines.

With a cylindrical Debye-Scherrer specimen, the intensity of the reflected beam is given by the expression

$$I = \text{constant} \cdot I_0 \tau \alpha p e^{\frac{-2B \sin^2 \theta}{\lambda^2}} \frac{1 + \cos^2 2\theta}{\cos^2 \theta \sin \theta} \{F(hkl)\}^2, \quad (6)$$

where

I = intensity of the reflected beam,

I_0 = intensity of the incident beam,

τ = the micro-absorption factor,

α = the macro-absorption factor,

p = the planar factor,

$e^{\frac{-2B \sin^2 \theta}{\lambda^2}}$ = the Debye-Waller temperature factor

$\frac{1 + \cos^2 2\theta}{\cos^2 \theta \sin \theta}$ = the angular factor, and

$F(hkl)$ = the structure factor.

If S is a parameter which varies from 1 for perfect order to 0 for total disorder, the intensities I_s of the superlattice lines and those of the main lattice lines, I_m , in Cu_3Au may be written

$$I_s = K I_0 \tau_s \alpha p_s e^{\frac{-2B \sin^2 \theta_s}{\lambda^2}} \frac{1 + \cos^2 \theta_s}{\cos^2 \theta_s \sin \theta_s} (f_{\text{Au}} - f_{\text{Cu}})^2 S^2, \quad (7)$$

$$I_m = K I_0 \tau_m \alpha p_m e^{\frac{-2B \sin^2 \theta_m}{\lambda^2}} \frac{1 + \cos^2 \theta_m}{\cos^2 \theta_m \sin^2 \theta_m} (f_{\text{Au}} + 3f_{\text{Cu}})^2. \quad (8)$$

Their ratio may therefore be written concisely as

$$\frac{I_s}{I_m} = C \cdot \phi(\theta) \cdot \psi(f_{\text{Au}}, f_{\text{Cu}}) \cdot S^2 \cdot \frac{\tau_s}{\tau_m} \quad (9)$$

or

$$\frac{\tau_s}{\tau_m} = \left(\frac{I_s}{I_m} \right) / C \cdot \phi(\theta) \cdot \psi(f_{\text{Au}}, f_{\text{Cu}}) \cdot S^2. \quad (10)$$

For a "faulted lattice" the R.H.S. of equation (10) will be very nearly unity, while if the conception of lattice domains is correct, it must vary with the experimental conditions. The simplest experiment to perform would be to take a given sample of known crystallite and "domain size" and compare the ratio τ_s/τ_m for different wave-lengths. The best "domain size" to take would be of the order 10^{-5} cm., for then the superlattice lines will be sharp enough to fix the background level accurately in the photometry, and at the same time they will be diffuse enough to ensure the existence of domains or faults. For the purposes of calculation, then, we shall take the crystallites to have a radius of 10^{-3} cm. and the domains to have a radius of 10^{-5} cm., sizes which would be very easily realizable in practice.

We compute the linear absorption coefficient for Cu_3Au from the expression

$$\mu_{\text{Cu}_3\text{Au}} = 12.3 \left\{ 0.4925 \left(\frac{\mu}{\rho} \right)_{\text{Cu}} + 0.5075 \left(\frac{\mu}{\rho} \right)_{\text{Au}} \right\}$$

$$= 6.05 \left(\frac{\mu}{\rho} \right)_{\text{Cu}} + 6.25 \left(\frac{\mu}{\rho} \right)_{\text{Au}},$$

where the value 12.3 is the density of Cu_3Au and 49.25, 50.75 are the weights per cent. of Cu and Au respectively in the alloy. Values of $\mu_{\text{Cu}_3\text{Au}}$ for different wave-lengths are given in Table IV.

We choose the main lattice line, 220, and its superlattice neighbour, 300, for obtaining the ratio τ_s/τ_m , for, since the lattice parameter of Cu_3Au is 3.74 Å., both these lines will be accommodated on the film for all radiations ranging from $\text{AgK}\alpha$ to $\text{CrK}\alpha$. The calculated values of τ_s , τ_m and their ratio are given in Table IV.*

TABLE IV.

Variation in the ratio τ_s/τ_m for the reflections 300, 220 from Cu_3Au with radiation of different wave-length, assuming domains of radius 10^{-5} and a crystallite radius 10^{-3} cm.

Radiation.	λ .	$\mu_{\text{Cu}_3\text{Au}}$.	Main line, 220.			Superlattice line, 300.			τ_s/τ_m .
			$\sin^2 \theta$.	μa .	$100\tau_m$.	$\sin^2 \theta$.	μa .	$100\tau_s$.	
Ag ..	0.560	576	0.045	0.576	42.8	0.050	0.00576	99	2.30
Rh ..	0.615	752	0.054	0.752	34.5	0.061	0.0075	99	2.87
Mo ..	0.710	1100	0.070	1.100	21.5	0.080	0.011	98.5	4.59
Cu ..	1.54	1658	0.340	1.658	14.0	0.381	0.017	97.5	6.97
Ni ...	1.66	2014	0.396	2.014	11.0	0.444	0.020	97.3	8.84
Fe ..	1.93	3035	0.530	3.035	7.5	0.600	0.030	96.0	12.8
Cr ...	2.29	4280	0.750	4.280	6.5	0.841	0.043	94.0	14.5

In Table V. we give the values of τ_s/τ_m assuming lattice faults. The ratio deviates slightly from unity with the longer wave-lengths on account of the variations of τ with Bragg angle as well as with μa .

The ratios τ_s/τ_m for the two theories are plotted in fig. 2. The break in the curve for τ_s/τ_m (domains) is caused by the presence of the gold L- and the copper K-absorption edges. The enormous difference between the τ_s/τ_m ratios should afford an easy experimental criterion for a choice between the two theories.

We have purposely chosen a ratio of "domain radius" to crystallite radius to give a large slope to the $(\tau_s/\tau_m, \lambda)$ -curve. If we make the domains larger, or the crystallites smaller, we shall, of course, diminish

* *Note added in Proof.*—In the case of tetragonal martensite, H. Lipson and Audrey M. B. Parker (J. Iron and Steel Inst., Nov. 1943, advance copy) have found a similar wave-length dependence for the intensities of lines within the group $\{hkl\}$. This is probably due to the shapes of the martensite crystals, which may be platy or needle-like, thus making the τ -factor appreciably smaller for X-rays travelling in the long direction of the crystal.

the slope of the curve until the ratio τ_s/τ_m approaches unity for all values of λ .

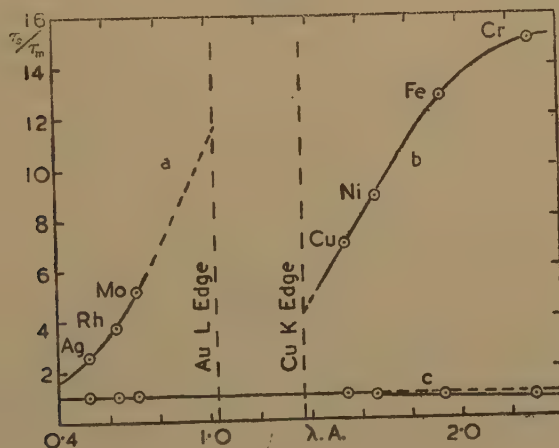
Assuming the domain theory to be correct, then the superlattice lines should always appear to be too strong and the degree of order

TABLE V.

The ratio τ_s/τ_m for the reflections 300, 220 from Cu_3Au with radiations of different wave-length, assuming lattice faults and a crystallite radius of 10^{-3} cm.

Radiation.	λ .	$\mu_{\text{Cu}_3\text{Au}}$.	Main line, 220.			Superlattice line, 300.			τ_s/τ_m .
			$\sin^2 \theta$.	μa .	$100\tau_m$.	$\sin^2 \theta$.	μa .	$100\tau_s$.	
Ag ..	0.560	576	0.045	0.576	42.8	0.050	0.576	42.8	1.00
Rh...	0.615	752	0.054	0.752	34.5	0.061	0.752	34.5	1.00
Mo...	0.710	1100	0.070	1.100	21.5	0.080	1.100	21.5	1.00
Cu ..	1.54	1658	0.340	1.658	14.0	0.381	1.658	14.1	0.99
Ni ...	1.66	2014	0.396	2.014	11.0	0.444	2.014	11.5	0.96
Fe ..	1.93	3035	0.530	3.035	7.5	0.600	3.035	8.3	0.90
Cr ...	2.29	4280	0.750	4.280	6.5	0.841	4.280	7.4	0.88

Fig. 2.



A comparison of the ratios τ_s/τ_m for the lines 300, 220 of Cu_3Au with different wave-lengths, made on the basis of (a) and (b) lattice domains of radius 10^{-5} cm, in crystallites of radius 10^{-3} cm., and (c) lattice faults in crystallites of radius 10^{-3} cm.

too high unless a τ -factor correction is made to the intensities of the spectra. This also applies to cases when the superlattice lines are sharp, since it is possible for the domains to be much smaller than the parent grains, while still being large enough to give sharp reflections. Thus

photographs may convey the impression of long-range order with a value of S greater than unity.

Conclusions.

The micro-absorption factor has been applied to the atomic scattering factors for filings of copper and nickel. It has been shown that grain size could account for the lower f values observed by Brindley and Spiers just as well as the theory based on random lattice displacements.

It has also been shown that by plotting the ratio of the micro-absorption factors for the lattice domains and crystallites, it is experimentally possible to verify the existence of ordered domains in the lattice of Cu_3Au or to prove that the broadening of the superlattice lines is caused by "faulting" analogous to that found in the cobalt lattice.

Acknowledgement.

The author thanks the English Electric Co., Ltd., for permission to publish this paper.

References.

- (1) A. Taylor, *Phil. Mag.* (In the press.)
- (2) A. J. Bradley, *Proc. Phys. Soc.* xlvii. p. 879 (1935).
- (3) G. W. Brindley and F. W. Spiers, *Phil. Mag.* xx. p. 865 (1935).
- (4) F. W. Jones and C. Sykes, *Proc. Roy. Soc.* clvii. p. 213 (1936); clxvi. p. 376 (1938).
- (5) O. S. Edwards and H. Lipson, *Proc. Roy. Soc.* clxxx. p. 268 (1942).

XLVIII. *Exercises on a Tigh trope.*

By E. H. NEVILLE*.

[Received April 12, 1944.]

IN the more familiar questions on a shallow catenary arc, the arc is rendered intrinsically determinate by the slack, the sag, or perhaps the maximum tension. Imagine, however, a cord passing over a number of smooth pegs in one vertical plane, or through a number of smooth rings, the total length of the cord being only slightly in excess of the sum of the linear distances from point to point; for example, a continuous loop thrown round three pegs forming a triangle whose perimeter is almost the length of the loop. Before we can calculate the form of any one festoon, we must know how the slack in the whole cord is shared between the several festoons. This is the problem with which this article is concerned. A process is developed for dealing with the general problem to any required degree of accuracy, and the first approximation

* Communicated by the Author.

to the distribution of the spare cord between the festoons is found to have a very simple form.

1. We require a lemma which we enunciate at once to avoid breaking the thread of the argument. Formal proof is superfluous.

If $C = \text{ch}(p \pm x)$, $S = \text{sh}(p \pm x)$, where x is a function of a variable t , successive derivatives with respect to t are given by

$$C^n = C\Phi^n(x) \pm S\Psi^n(x), \quad S^n = S\Phi^n(x) \pm C\Psi^n(x),$$

$$\text{where} \quad \Phi^{n+1}(x) = \dot{\Phi}^n(x) + \dot{x}\Psi^n(x), \quad \Psi^{n+1}(x) = \dot{\Psi}^n(x) + \dot{x}\Phi^n(x),$$

$$\text{with} \quad \Phi^0(x) \equiv 1, \quad \Psi^0(x) \equiv 0.$$

The auxiliary functions, as far as the sixth order, are explicitly

n	0	1	2	3	4	5
$\Phi^n(x)$	1	0	\dot{x}^2	$3\dot{x}\ddot{x}$	$4\dot{x}\ddot{x} + 3\ddot{x}^2 + \dot{x}^4$	$5\dot{x}x^{(4)} + 10\dot{x}\ddot{x} + 10\dot{x}^3\ddot{x}$
$\Psi^n(x)$	0	\dot{x}	\ddot{x}	$\ddot{x} + \dot{x}^3$	$x^{(4)} + 6\dot{x}^2\ddot{x}$	$x^{(5)} + 10\dot{x}^2\ddot{x} + 15\dot{x}\ddot{x}^2 + \dot{x}^5$
		n			6	
			$\Phi^n(x)$	$6\dot{x}x^{(5)} + 15\ddot{x}x^{(4)} + 10\ddot{x}^2 + 20\dot{x}^3\ddot{x} + 45\dot{x}^2\ddot{x}^2 + \dot{x}^6$		
			$\Psi^n(x)$	$x^{(6)} + 15\dot{x}^2x^{(4)} + 60\dot{x}\ddot{x}\ddot{x} + 15\ddot{x}^3 + 15\dot{x}^4\ddot{x}$		

If we write

$$\Phi^{n+2}(x) = (n+2)\dot{x}x^{(n+1)} + \phi^n(x), \quad \Psi^{n+2}(x) = x^{(n+2)} + \psi^n(x),$$

then $\phi^n(x)$, $\psi^n(x)$, except in the trivial case of $\phi^0(x)$, involve no derivatives of order higher than n ; for $n > 2$, the expressions

$$C\phi^{n-2}(x) \pm S\psi^{n-2}(x), \quad S\phi^{n-2}(x) \pm C\psi^{n-2}(x)$$

are the expressions for C^n , S^n with the terms involving $x^{(n)}$ and $x^{(n-1)}$ omitted. The functions $\phi^n(x)$, $\psi^n(x)$ have their own recurrences:

$$\phi^{n+1}(x) = \dot{\phi}^n(x) + \dot{x}\psi^n(x) + (n+2)\ddot{x}x^{(n+1)},$$

$$\psi^{n+1}(x) = \dot{\psi}^n(x) + \dot{x}\phi^n(x) + (n+2)\dot{x}^2x^{(n+1)},$$

with

$$\phi^0(x) = -\dot{x}^2, \quad \psi^0(x) \equiv 0.$$

2. If two festoons meet at a peg, it is the actual tension, or in geometrical terms it is the height above the directrix, that has the same value for the two arcs at that point. By hypothesis the height is large. We therefore use the reciprocal of the height as independent variable, and our first problem is as follows:—

Given the slope B of a chord of unit length in a tight catenary, and the height $1/t$ of the upper end of the chord above the directrix, develop the length s of the arc in powers of t .

If c is the parameter of the catenary and x , $x - \cos B$ are the abscissæ of the upper and lower ends, we have

$$\frac{1}{t} = c \operatorname{ch} \frac{x}{c}, \quad \frac{1}{t} - \sin B = c \operatorname{ch} \frac{x - \cos B}{c};$$

these two equations implicitly determine c and x , and s follows as

$$c \operatorname{sh} \frac{x}{c} - c \operatorname{sh} \frac{x - \cos B}{c}.$$

The slopes at the two ends of the arc are $\operatorname{sh}(x/c)$, $\operatorname{sh}\{(x - \cos B)/c\}$, and these differ but slightly from $\tan B$, the one in excess, the other in defect. We therefore introduce a constant q by the Gudermannian relation

$$\operatorname{sh} q = \tan B, \quad \operatorname{ch} q = \sec B,$$

and writing $x/c = q + u$, $(x - \cos B)/c = q - v$,

we have for the determination of u and v as functions of t the pair of equations

$$v + u = t \cos B \operatorname{ch}(q + u), \quad \operatorname{ch}(q - v) = (1 - t \sin B) \operatorname{ch}(q + u),$$

with initial values $v_0 = 0$, $u_0 = 0$.

To solve this pair of equations, write

$$\begin{aligned} \operatorname{ch}(q + u)/\operatorname{ch} q &= C, & \operatorname{sh}(q + u)/\operatorname{ch} q &= S, \\ \operatorname{ch}(q - v)/\operatorname{ch} q &= D, & \operatorname{sh}(q - v)/\operatorname{ch} q &= T; \end{aligned}$$

then the equations become

$$v + u = tC, \quad C - D = k t C,$$

where $k = \sin B$, and initially $C = D = 1$, $S = T = k$. The initial values of the successive derivatives of the four functions C , S , D , T are given, in the notation of the lemma, by

$$\begin{aligned} C_0^n &= \Phi_0^n(u) + k \Psi_0^n(u), & S_0^n &= k \Phi_0^n(u) + \Psi_0^n(u), \\ D_0^n &= \Phi_0^n(v) - k \Psi_0^n(v), & T_0^n &= k \Phi_0^n(v) - \Psi_0^n(v), \end{aligned}$$

in terms of the initial values of the four auxiliary functions. To utilize these formulæ we must know the initial values of the successive derivatives of u and v . One relation between $u_0^{(n)}$ and $v_0^{(n)}$ is immediate; for all values of t ,

$$v^{(n)} + u^{(n)} = n C^{n-1} + t C^n,$$

and therefore

$$v_0^{(n)} + u_0^{(n)} = n C_0^{n-1}.$$

For a second relation between $u_0^{(n)}$ and $v_0^{(n)}$ we must look to the relation between C and D . If we differentiate only n times before putting $t = 0$, we have

$$\Phi_0^n(u) - \Phi_0^n(v) + k \{ \Psi_0^n(u) + \Psi_0^n(v) \} = k n C_0^{n-1},$$

and since $u_0^{(n)}$, $v_0^{(n)}$ occur in this equation only in the terms $\Psi_0^n(u)$, $\Psi_0^n(v)$, this is not a second relation between these initial values. But when we reconcile the two values of $v_0^{(n)} + u_0^{(n)}$ we have a relation between the initial values of derivatives of lower orders. Explicitly, differentiating $n+1$ times and expressing the functions Φ^{n+1} , Ψ^{n+1} in terms of ϕ^{n+1} , ψ^{n+1} , we have

$$(n+1)(\dot{v}_0 v_0^{(n)} - \dot{u}_0 u_0^{(n)}) = \{ \phi_0^{n-1}(u) - \phi_0^{n-1}(v) \} + k \{ \psi_0^{n-1}(u) + \psi_0^{n-1}(v) \}.$$

and to see that this is a new relation between $u_0^{(n)}$ and $v_0^{(n)}$ we have only to find the values of \dot{u}_0 and \dot{v}_0 . We have

$$\dot{v}_0 + \dot{u}_0 = 1;$$

TABLE OF INITIAL VALUES.

$n.$	1.	2.	3.	4.	5.
$u^{(n)}$	$\frac{1}{2}$	$\frac{5k}{12}$	$\frac{3+3k^2}{8}$	$\frac{487k+15k^3}{240}$	$\frac{195+730k^3-165k^4}{96}$
$v^{(n)}$	$\frac{1}{2}$	$\frac{7k}{12}$	$\frac{3+7k^2}{8}$	$\frac{593k+345k^3}{240}$	$\frac{195+1154k^2+195k^4}{96}$
$\Phi^n(u)$	0	$\frac{1}{4}$	$\frac{5k}{8}$	$\frac{39+61k^3}{48}$	$\frac{687k+165k^3}{96}$
$\Phi^n(v)$	0	$\frac{1}{4}$	$\frac{7k}{8}$	$\frac{39+133k^2}{48}$	$\frac{873k+835k^3}{96}$
$\Psi^n(u)$	$\frac{1}{2}$	$\frac{5k}{12}$	$\frac{4+3k^2}{8}$	$\frac{637k+15k^3}{240}$	$\frac{288+945k^2-165k^4}{96}$
$\Psi^n(v)$	$\frac{1}{2}$	$\frac{7k}{12}$	$\frac{4+7k^2}{8}$	$\frac{803k+345k^3}{240}$	$\frac{288+1609k^2+195k^4}{96}$
C^n	$\frac{k}{2}$	$\frac{3+5k^2}{12}$	$\frac{9k+3k^3}{8}$	$\frac{65+314k^2+5k^4}{80}$	$\frac{975k+1110k^3-165k^5}{96}$
D^n	$-\frac{k}{2}$	$\frac{3-7k^2}{12}$	$\frac{3k-7k^3}{8}$	$\frac{65-46k^2-115k^4}{80}$	$\frac{585k-774k^3-195k^5}{96}$
S^n	$\frac{1}{2}$	$\frac{2k}{3}$	$\frac{1+2k^2}{2}$	$\frac{52k+20k^5}{15}$	$\frac{9+51k^2}{3}$
T^n	$-\frac{1}{2}$	$-\frac{k}{3}$	$-\frac{1}{2}$	$-\frac{38k-20k^3}{15}$	$-\frac{9+23k^2-20k^4}{3}$
$\phi^n(u)$	0	$\frac{3+25k^2}{48}$	$\frac{100k+75k^3}{48}$	$\frac{453+3530k^2+345k^4}{192}$	$\frac{17787k+30604k^3-4095k^5}{288}$
$\phi^n(v)$	0	$\frac{3+49k^2}{48}$	$\frac{140k+245k^3}{48}$	$\frac{453+6566k^2+3885k^4}{192}$	$\frac{23037k+82166k^3+19845k^5}{288}$
$\psi^n(u)$	$\frac{1}{8}$	$\frac{5k}{8}$	$\frac{93+215k^2}{96}$	$\frac{1827k+865k^3}{144}$	$\frac{6303+39970k^2+1575k^4}{384}$
$\psi^n(v)$	$\frac{1}{8}$	$\frac{7k}{8}$	$\frac{93+455k^2}{96}$	$\frac{2358k+3410k^3}{144}$	$\frac{6303+69706k^2+43295k^4}{384}$
$\frac{v^{(n+1)}+u^{(n+1)}}{2}$	$\frac{6k}{12}$	$\frac{3+5k^2}{8}$	$\frac{540k+180k^3}{240}$	$\frac{195+942k^2+15k^4}{96}$	$\frac{122850k+139860k^3-20790k^5}{4032}$
$\frac{v^{(n+1)}-u^{(n+1)}}{2}$	$\frac{k}{12}$	$\frac{2k^2}{8}$	$\frac{53k+165k^3}{240}$	$\frac{212k^2+180k^4}{96}$	$\frac{8409k+61390k^3+19425k^5}{4032}$

the second relation is anomalous, since $\phi^0(x)$ does involve \dot{x} , but the relation reduces simply to

$$v_0^2 - \dot{u}_0^2 = 0,$$

and the result is unambiguous :

$$\dot{u}_0 = \frac{1}{2}, \quad \dot{v}_0 = \frac{1}{2}.$$

Thus for all values of n we have

$$\begin{aligned} \frac{1}{2}(n+1)(v_0^{(n)} + u_0^{(n)}) &= \frac{1}{2}n(n+1)C_0^n - 1, \\ \frac{1}{2}(n+1)(v_0^{(n)} - u_0^{(n)}) &= \{\phi_0^{n-1}(u) - \phi_0^{n-1}(v)\} + k\{\psi_0^{n-1}(u) + \psi_0^{n-1}(v)\}, \end{aligned}$$

and for $n > 1$ these formulæ determine the initial values of $u^{(n)}$ and $v^{(n)}$ recurrently by means of the initial values of derivatives of lower orders.

We are now in a position to tabulate systematically : results are given as far as $n=5$. For the arc s we have $s=c(S-T)\text{ch}q$, and therefore

$$Cs = (S-T)/t,$$

a formula from which the coefficients in the power series for s can be evaluated one by one.

Although the coefficient of t^n in $(S-T)/t$ is $(S_0^{n+1} - T_0^{n+1})/\overline{n+1}!$, derivatives of order $n+1$ appear in this coefficient only in the combination $u_0^{(n+1)} + v_0^{(n+1)}$, and since this is $(n+1)C_0^n$, the development of s to order n requires only initial values as far as order n . Moreover, to move the origin from the upper to the lower end of the arc is to use $1/t - k$ in the place of $1/t$, and therefore the substitution of $t(1+kt)^{-1}$ for t is equivalent to a change in the sign of k ; this functional property alone yields the coefficient of any odd power of t in s when all the preceding coefficients are known.

To the fifth order in t ,

$$\begin{aligned} s = 1 + \frac{1}{24}t^2 \cos^2 B \left\{ 1 + t \sin B + \frac{1}{80}t^2(21 + 55 \sin^2 B) \right. \\ \left. + \frac{1}{40}t^3 \sin B (21 + 15 \sin^2 B) \right\}. \end{aligned}$$

If the chord is of length b , the relation between the length s of the arc and the depth $1/t$ of the directrix is obtained by replacing s by s/b and t by bt throughout. Thus if $s=b+\beta$, so that β is the slack in the rope, and if h, k are the horizontal and vertical projections of the chord, h being measured downwards from the same end as the distance $1/t$, we have, again to the fifth order in t ,

$$24\beta = bh^2t^2 \left\{ 1 + kt + \frac{21b^2 + 55k^2}{80}t^2 + \frac{21b^2k + 15k^3}{40}t^3 \right\}.$$

3. Suppose now that two festoons meet at a smooth peg. Let the chords be a , with horizontal and vertical projections f, g , and b , with horizontal and vertical projections h, k , and let the total length of the rope be $a+b+\sigma$. Then, if the lengths of slack in the festoons are α, β , we have

$$24\alpha = af^2t^2(1 + g^{(1)}t + g^{(2)}t^2 + \dots), \quad 24\beta = bh^2t^2(1 + k^{(1)}t + k^{(2)}t^2 + \dots),$$

where

$$\begin{aligned} g^{(1)} &= g, & g^{(2)} &= (21a^2 + 55g^2)/80, \dots, \\ k^{(1)} &= k, & k^{(2)} &= (21b^2 + 55k^2)/80, \dots, \end{aligned}$$

these coefficients all having known constant values, and t having the same value in the two expansions. Since $\alpha + \beta = \sigma$, the common value of t is given to any arbitrary order of accuracy by the equation

$$24\sigma = (af^2 + bh^2)t^2 + (af^2g^{(1)} + bh^2k^{(1)})t^3 + (af^2g^{(2)} + bh^2k^{(2)})t^4 + \dots$$

The form of this equation is enough to show that the slack σ is not itself a suitable independent variable.

If $\{\sigma/(a+b)\}^{3/2}$ is negligible, then powers of t above the second are negligible, and we have simply

$$\frac{\alpha}{af^2} = \frac{\beta}{bh^2} = \frac{\sigma}{af^2 + bh^2}.$$

This result can be extended immediately to any number of festoons:

If a uniform rope that is almost tight hangs in festoons through any number of smooth rings, then to a first approximation the ratio of the length of slack in any festoon to the chord of the festoon is proportional to the square of the horizontal span of the festoon.

As a particular case,

If a uniform tightrope passes through any number of smooth rings which are all collinear, or all in the same horizontal plane, then to a first approximation the length of slack in a festoon is proportional to the cube of the span of that festoon.

To go beyond the first approximation when there are more than two festoons involves changes of origin, or, better, the introduction of other variables. For example, we can deal with four festoons by means of two variables t_1, t_2 connected with each other by the equation

$$1/t_1 - k_1 = 1/t_2 - f_2,$$

that is,

$$t_1 - t_2 = (f_2 - k_1)t_1t_2,$$

and with the total slack σ by the equation

$$24\sigma = \{(a_1f_1^2 + b_1h_1^2)t_1^2 + (a_2f_2^2 + b_2h_2^2)t_2^2\} \\ + \{(a_1f_1^2g_1^{(1)} + b_1h_1^2k_1^{(1)})t_1^3 + (a_2f_2^2g_2^{(1)} + b_2h_2^2k_2^{(1)})t_2^3\} + \dots$$

It is a simple matter to substitute $t_1\{1 + (f_2 - k_1)t_1\}^{-1}$ for t_2 to any desired order and so to obtain an unsymmetrical equation in t_1 alone, but in practice we shall probably prefer to retain two variables and a pair of simultaneous equations. In any case the first approximation is provided by

$$t_1 = t_2 = \sqrt{\{24\sigma/(a_1f_1^2 + b_1h_1^2 + a_2f_2^2 + b_2h_2^2)\}},$$

and in all questions reduced to functional equations the determination of an effective first approximation is half the battle.

XLIX. *Dimensions.*

By Professor W. WILSON, F.R.S.*

[Received January 1, 1944.]

IN a recent further communication on this subject † Professor Dingle finds it impossible to imagine how I could have failed to realize that I had reached an absurdity. Referring to something I recently wrote ‡ he says: "How he can assert his equation (15), adopting, as he does, my dimensions for the magnitudes involved, or write $[M]=[T^3]$ after M, L and T have been chosen as fundamental magnitudes, and proceed without realizing that he has reached an absurdity, I cannot imagine." But *I did realize that I had reached an absurdity*, indeed, several absurdities. My equation (15)—which by the way is not *my* equation, but a generally accepted one, firmly based on experimental deliverances, which I made use of and numbered (15)—is not responsible, as I will show. The absurdities can easily be traced to Dingle's own premisses, *which I did not just simply adopt, but adopted for the sake of argument and for the deliberate purpose of demonstrating that they led to absurdities*. Let us start with his equation (16) in his latest paper. It is

$$[V]=[\alpha E']. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

In this statement, as Dingle tells us, V means potential difference and E' means electromotive force. *Both* of these quantities can be, and frequently or usually are, measured by a voltmeter in terms of *the same unit*, the volt, or one joule per coulomb. Indeed, Dingle himself has laid down, in a recent book on Sub-Atomic Physics, that the electromotive force of a cell is identical with the potential difference between its poles in the limiting case for which the current vanishes. We *must* therefore infer that

$$[\text{volts}]=[\alpha \text{ volts}]$$

and, consequently, that α is without dimensions. Now Dingle's (15), in this paper, is

$$[\alpha]=[LTF^{-1}]=[M^{-1}T^3], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and consequently $[M^{-1}T^3]$ is without dimensions. So we get

$$[M]=[T^3]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Professor Dingle declares this to be an absurdity and he is undoubtedly right.

* Communicated by the Author.

† Phil. Mag. vol. xxxiv. p. 588 (1943).

‡ Phil. Mag. vol. xxxiii. p. 631 (1942).

In order to make it even more convincing that this, and indeed other, absurdities follow from the adoption of Dingle's dimensions, let us take his equation (18), which is

$$[\beta] = [FL^{-2}] = [ML^{-1}T^{-2}], \quad (4)$$

The quantity β appears to be a dielectric constant. At any rate, according to Dingle, it has the same dimensions as a dielectric constant and I may therefore be permitted to write

$$[K] = [ML^{-1}T^{-2}], \quad (5)$$

Similarly we can infer from Dingle's (19) and (20) that

$$[\mu] = [ML^{-1}T^{-2}], \quad (6)$$

μ being a permeability. Combining (5) and (6) we get

$$[\mu^{-\frac{1}{2}}K^{-\frac{1}{2}}] = [M^{-1}LT^2].$$

But Dingle's (26) assures us that

$$[\mu^{-\frac{1}{2}}K^{-\frac{1}{2}}] = [LT^{-1}].$$

Hence

$$[LT^{-1}] = [M^{-1}LT^2],$$

and once again we reach the absurdity

$$[M] = [T^3].$$

Professor Dingle claims that a factor α , with the dimensions $[M^{-1}T^3]$, is necessary in his equation (14), which is

$$i = \alpha q/t, \quad (7)$$

in order to make it homogeneous, *i. e.* in order that both sides may have the same dimensions. We have, however, deduced, from Dingle's own statements, that α is without dimensions. Equation (7) therefore becomes

$$i = q/t \times (\text{some number}), \quad (7a)$$

and the number may be conveniently chosen to be unity, as indeed everybody, except Professor Dingle, does choose it. Indeed, one may say that this quantity α , which Dingle still claims has some sort of correspondence with permeability, has vanished into thin air.

How is it, one may inquire, that the absurdity (this is Dingle's name for it), expressed in (3), emerges so consistently? The question is easily answered. It can be traced to incorrect inferences from certain definitions. In particular there is his definition (i) of Quantity of Electricity, which makes the quantity, Q , proportional to the force it exerts on an arbitrary standard charge *in vacuo* and at a fixed distance away, *i. e.*

$$Q = NF, \quad (8)$$

where N is the constant of proportionality. There is nothing seriously wrong with this definition. I have used it myself—with a little more precision in its statement—in vol. ii. of my 'Theoretical Physics,' published ten years ago. What *is* wrong is the inference

$$[Q] = [F] = [MLT^{-2}]$$

Consequently

$$[\gamma M] = [L^3 T^{-2}],$$

and

$$[c^2] = [L^2 T^{-2}]$$

(c is, of course, the velocity of light in free space). Hence

$$[M'] = \left[\frac{\gamma M}{c^2} \right] = [L].$$

M' is often termed (by relativists) the *mass* of the particle whose mass (*e. g.* in grammes) is M . Undoubtedly M' has the dimensions of a length and, if there is any "nonsense" associated with this, it must lie in the description of M' as a *mass*. But surely this description is quite appropriate, since M' is simply the product of a mass, in the usual sense of the term, and the *universal constant* γ/c^2 . Indeed M' might be used everywhere in Physics as a substitute for M without changing the form, or significance, of a single equation. The only change one would notice would be the replacement of γ/c^2 by unity.

It is of interest to make a rough estimate of the M' of the sun. Approximately we have

$$M = 2 \times 10^{33} \text{ grammes,}$$

$$\gamma = 6.7 \times 10^{-8} \text{ gr.}^{-1} \text{ cm.}^3 \text{ sec.}^{-2},$$

$$c^2 = 9 \times 10^{20} \text{ cm.}^2 \text{ sec.}^{-2}.$$

Hence

$$M' = \frac{6.7 \times 2 \times 10^{-8} \times 10^{33}}{9 \times 10^{20}} \text{ cm.,}$$

or, finally,

$$M' = 1.49 \text{ kilometres.}$$

Eddington, in his 'Mathematical Theory of Relativity,' assumes M' to be 1.47 kilometres. Dingle writes: "This is incidentally like my nonsense," *i. e.* Dingle's nonsense, "in speaking of my μ as equivalent to permeability." It is, of course, nothing of the kind, since Dingle's nonsense, as he himself has recognized, can be identified with an error.

The last thing, in Dingle's paper, that I wish to comment on, is his reference to the fine structure constant, the purely numerical character of which, by the way, becomes obvious when we remember that it signifies the ratio of two velocities. Dingle writes it in the form $ch/2\pi e^2$ and jumps at once to the conclusion that it *cannot* be dimensionless! Perhaps it will not be presumptuous on my part to give some information about this constant. It appears first in a paper by Arnold Sommerfeld, published in 1916, or soon after*. It emerged from the

* Bohr sent a single reprint of this paper, during the last war, to J. W. Nicholson, then Professor of Mathematics in King's College, London. Nicholson had a copy (or perhaps more than one copy) made of it, which he handed to me, drawing attention to the fact (as Bohr had also done) that Sommerfeld adopted in his work the same quantum conditions as those I had enunciated somewhat earlier (Phil. Mag. June (1915) and Feb. (1916)). These were, of course, the conditions

$$\oint p_i dq_i = n_i h.$$

application of the quantum conditions (given in the footnote) and appears in Sommerfeld's writings as the reciprocal, $2\pi e^2/c\hbar$, of the number which Dingle and others now term "fine structure constant." It can be identified with the ratio v_1/c , where v_1 is the velocity of the electron (of a hydrogen atom) in the first Bohr (circular) orbit. Indeed, it is quite easy to deduce

$$v_1 = 2\pi e^2 / \hbar, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

or, to be quite precise,

[illegible]

where ϵ is the number $M/(M+m)$, M and m being the masses of the nucleus and the electron respectively. All we have to do is to apply the appropriate quantum condition, which, in this case, amounts to the statement

$$2\pi \times (\text{Angular Momentum}) = h.$$

Now it should be noted that the dielectric constant (Dingle's κ or β) is, in fact, present in the denominator of (13), or (13*a*). The units used are electrostatic and, as the electron is supposed to be moving in free space, the value of κ is unity. Nobody usually troubles to use an explicit symbol for a multiplying factor when its value happens to be unity, and consequently the presence of κ is hidden from those who do not know how (13) is derived. With other units (13) would appear in the form

$$v_1 = 2\pi e^2 / \kappa \hbar,$$

or, to be quite exact, in the form (13 a), since indeed it originates from

$$\frac{mv_1^2}{r} = \frac{e^2}{\kappa r^2}$$

Or, strictly,

$$mv_1^2/r = \epsilon^2 e^2 / \kappa r^2,$$

where ϵ is the numerical factor already defined.

Professor Dingle is therefore (no doubt inadvertently) misrepresenting the fine structure constant when he assumes that it does not contain κ , and he is misrepresenting it still more when he contends that the introduction of κ —which of course would be necessary to make it dimensionless if one were under the delusion that it is not already there—“is in fact a fundamental change . . .” It is merely making explicit, by the symbol κ , a quantity which is in the formula $ch/2\pi e^2$ all the time, but hidden under the disguise of unity.*

* *Vide* A. Sommerfeld : 'Atombau und Spektrallinien,' 5th edition (1931), vol. i. §§ 3 and 4, and especially p. 94.

L. Notices respecting New Books.

A Survey of General and Applied Rheology. By G. W. SCOTT-BLAIR. (Pitman and Sons, Ltd.) Price 18s.

DR. SCOTT-BLAIR has probably done more than any man to make British scientists rheologically minded and a work from his pen is always worth reading. The present memoir was intended for a course of lectures on rheology which he was invited to give at a United States university but which, owing to the war, has not yet been read. The lectures in book form are attractive to read, infused as they are with the author's personality; indeed, they are much more an expression of his own views and experience than was the first book. It is evident that an 'Introduction to Industrial Rheology' roused much interest, since a not inconsiderable proportion of the subject-matter of the new book arises out of correspondence on the first. Opportunity has been taken to put the reader *au fait* with work in rheology to the present date. At the same time, the new book in no way repeats the old. The later chapters on psychological aspects of rheology cover the author's recent work and introduce the reader to what must be for many a new angle on the science.

E. G. R.

Experiment and Theory in Physics. By MAX BORN. [Pp. 44.] (Cambridge: The University Press, 1943.) Price 2s. net.

IN this expansion of an address which he gave to the Durham Philosophical Society and the Pure Science Society, King's College, at Newcastle-on-Tyne, Professor Born is concerned with the relative places of theory and experiment in physical investigation. In particular, he is suspicious of the recent theories of Eddington and Milne, which seem to prove more than could reasonably be expected from the few general postulates on which they appear to be based, and this has led him to examine how, in fact, advances in physics do come about. He finds that it is neither from theory nor experiment, working alone, but that in the march of progress, theory and experiment move forward alternately, each movement being made possible in general only by the previous advance of the other, just as in marching, one foot can only advance when the other has done so.

For example, electrodynamics could be built up from the postulate that there is to be a function, the *action* built up from E and H , which is to have a stationary variation, that the equations are to be linear in E and H and that they are to be invariant under Lorentz transformation. But, in fact, if we did not already know E and H from experiment, there is no reason to suppose that the step would be taken, or, at any rate, that the results would be identified with the phenomena of the electric field. The situation with regard to the latest form of Eddington's theory is somewhat similar: the results are obtained by means of a particular non-commutative algebra, the E -numbers, But, Born asks, even if this particular algebra had been studied long ago what reason would there have been for applying it to general field theory? It is only because *experiment* calls for it, that it is used.

The lecturer's objections to Milne's theory are of a different nature. He denies that Milne's imaginary observations, such as reflecting terrestrial light from a distant nebula, are legitimate conceptual expansions of the observations which we do make.

This very short account of the pamphlet will no doubt fail to give an accurate picture of the ideas which Professor Born has tried to convey. But if it shows those interested in the matter that they ought to read the original, it will have served its proper purpose.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

LI. *A General Theory of Antennæ.*

By J. AHARONI, Ph.D.†

[Received October 14, 1943.]

SUMMARY.

A general theory for calculating the transmitting and receiving properties of antennæ is outlined. It contains Hallén's theory of antennæ which consist of thin wires as a special case. As far as these are concerned it may be regarded, therefore, as a different derivation of his one-dimensional equations. The present discussion applies to any form of conductor.

I. INTRODUCTION.

THE relative simplicity of the ordinary circuit theory results from two main assumptions which are perfectly legitimate in systems which are small compared with the wave-length :

1. The radiation can be neglected. This implies that the currents are quasi-stationary and the charges which accumulate quasi-static. The magnetic and electric fields vanish at infinity more rapidly than r^{-1} ($r \rightarrow \infty$), so that an integration of Poynting's vector over an infinite sphere gives zero and no radiation results.

2. The system consist of meshes, in each of which a current is circulating. These currents are at any moment constant along the wire which forms the mesh, and charges can accumulate only over the plates of condensers which may interrupt the continuity of the wires. Thus the geometrical shape of the electric and magnetic fields which are contributed by each mesh can be regarded as known and is used to calculate certain parameters : inductances, capacitances, and resistances.

By a well-known method certain differential equations can be obtained in special cases, when the current varies along the wire. Thus for a transmission line, which is regarded as a succession of very small circuits, the telegraphic equations are obtained, which contain as parameters the inductances, capacitances, resistances, and leakage losses per unit length. These are calculated from infinitely long lines, so that the theory applies only as long as the spacing of the line is small compared with its length.

For systems which radiate electromagnetic energy no such general theory of the type of the circuit theory has been developed so far. In antennæ the radiated energy represents the main part of the work of

† Communicated by Professor H. T. Flint.

the E.M.F. and far exceeds the heat losses. Apart from that, the current is not uniform along the conductor and meshes cannot be introduced. Any attempt to generalize the theory of transmission lines must be regarded as futile or at least as highly artificial.

The theory of radiating and receiving systems is, of course, governed by Maxwell's equations, but unfortunately the direct integration of these equations, even in the simplest cases of practical importance, lead to great mathematical difficulties†. It is also difficult to obtain in this way a link with the ordinary circuit theory, which consists of certain differential equations for the currents only, the field intensities not appearing directly. Obviously we cannot introduce circulating currents when dealing with antennæ, and these will have to be replaced in general by the current density. We may well regard a theory which contains the current-density as the only unknown as a generalization of the circuit theory, indeed, as the exact circuit theory, as it must represent an exact consequence of Maxwell's equations, with nothing neglected.

An important contribution towards this aim was made by Hallén. His theory applies to thin wires only the diameter of which is negligible when compared with their length; $\left(2 \log \frac{2l}{r_0}\right)^{-1}$ ($2l$ length, $2r_0$ diameter of wire) is nevertheless treated as a finite quantity. In this case we can still speak of a total current at each point of the wire, so that the current is a function of one variable only, the distance along the wire. Hallén shows, by introducing a number of approximations of "one dimensional accuracy" in the above sense, that the current obeys a certain integral equation. He applies his theory to a straight wire antenna and to a circular loop. The physical significance of the integral equation is, however, not brought to light in Hallén's derivation.

In the following a new approach towards the solution of antenna problems is made and a far more general result than was achieved by Hallén is reached. It is not restricted to thin wires and contains Hallén's theory as a special case.

The question to be answered is: can a method of calculating the electrical current be developed from Maxwell's equations, by making use, as in the circuit theory, of some knowledge about the current, something corresponding to assumption (2) mentioned before. Assumption (1) must be dropped, of course. It will be shown that if the lines of flow are stationary (or can be assumed so to a high degree of approximation) a general answer can be given to the above question. The density of the electric current obeys an integral equation,

$$\frac{|\mathbf{J}_e|}{\sigma} + \iiint K(\mathbf{x}, \mathbf{y}, \mathbf{z}; \mathbf{x}' \mathbf{y}' \mathbf{z}') |\mathbf{J}_e(\mathbf{x}' \mathbf{y}' \mathbf{z}')| d\mathbf{x}' d\mathbf{y}' d\mathbf{z}' = |\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})|. \quad (1)$$

† A direct integration of Maxwell's equations to obtain the forced oscillations of a cylindrical antenna was carried out by L. V. King. We shall refer to his paper in § VIII.

$|\mathbf{J}_e|$ is the amplitude of the density of the electric current. $K(x, y, z; x'y'z')$ is a certain "impedance kernel" which is easily obtained if the line along which the current flows at each point is known. $|\mathbf{F}(x, y, z)|$ is the amplitude of the oscillatory E.M.F. It is assumed to be seated in the conductor and represents the generator. At present we have in mind a transmitting antenna. With a receiving antenna we shall deal later. σ is the conductivity. In the ordinary circuit theory we need not make any assumption about the distribution of the E.M.F., only its total line integral appears in the circuit equations. When dealing with antennæ a more detailed model must be given. We shall in most cases assume \mathbf{F} to be acting in a very small region, yet yielding a finite volume integral (or a line integral in the case of thin wires). It will thus represent a point generator the dimensions of which are negligible when compared with those of the antenna. It may be added here that \mathbf{F} may also represent a power absorbing force, instead of power generating one. It will then represent the receiver which is attached to a receiving antenna. We shall deal with this aspect of the problem in § VII.

To be sure, the integral equation (1) is an exact consequence of Maxwell's equations, and expresses a hitherto unknown relation which is obeyed by the current density. In the sense explained above it represents the exact theory, as neither the ohmic nor radiation losses are neglected. To apply it we must determine the kernel. This can be done easily in all cases in which stationary lines of flow exist and are known *a priori*. In many instances this is the case. Particularly in thin wires the lines of flow may be assumed as parallel to the central line of the wire. But there exist many other cases where, from the symmetry, the lines of flow are known, especially if the frequency is high and the current is mainly a surface current.

In the same way as in the circuit theory, where the parameters L, C, R must be known, so here the kernel must be predetermined. The calculation of the ordinary circuit parameters is usually tedious, whereas in all cases in which the lines of flow are known the kernel can be written without any difficulty. But in the ordinary circuit theory, the currents obey a system of simple differential equations, the steady state solution of which is easily determined, whereas to solve an integral equation is still a formidable task. Nevertheless, it is easier to deal with a single integral equation than with a system of partial differential equations with attached boundary conditions.

As will be seen (1) is an integral equation of the second kind. But usually the conductivity is so high, that $\frac{|\mathbf{J}_e|}{\sigma}$ can be neglected, when the integral equation becomes of the first kind. This in itself is rather a disadvantage, but a great conductivity often makes it possible to get a more accurate picture of the lines along which the current flows and both a simpler and more accurate impedance kernel can be used.

The method of obtaining (1) consists mainly in the following steps:

(a) The derivation of a circuit theory which does not neglect the

radiation, for networks which are small compared with the wave-length. These networks can consist either of a very small loop, or of a short and thin piece of wire. For our ultimate purpose we shall be interested only in the latter, but it will be instructive to consider also a small loop. In small networks like these we can speak of the total current in each part, and even regard the current as constant along the loop or the wire. This will, at any rate, be exactly the case in infinitesimal components, which we shall consider finally. As will be shown, a circuit theory very similar to the ordinary one is obtained, the only difference is in the inductances and capacitances not being real as usual, but complex. Thus the impedance $j\omega L = j\omega(L_r + jL_i)$ contains a real component $-\omega L_i$ which is connected with the radiated power. The same applies to the capacitance.

(b) The calculation of the mutual effect between two short and thin pieces of wire the distance between which is not restricted. Here again we obtain mutual inductances and capacitances of complex value, but otherwise the result is quite similar to that of the ordinary circuit theory. For n small components (short pieces of wire or loops) we get n linear equations for the currents.

(c) The final step is to regard any conductor as consisting of a great number of small line elements in the direction of current flow. The greater the subdivision the greater the number of the linear equation, until in the limit the system of equations becomes equivalent to an integral equation.

II. THE DERIVATION OF THE ORDINARY CIRCUIT THEORY FROM MAXWELL'S EQUATIONS.

Before we proceed with the three steps mentioned in the introduction, it will be useful to describe briefly how the ordinary circuit theory can be derived from Maxwell's equations. Let us write these in their Gaussian form :

$$c \operatorname{curl} \mathbf{H} = 4\pi\sigma(\mathbf{E} + \mathbf{F}) + \epsilon \dot{\mathbf{E}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$c \operatorname{curl} \mathbf{E} = -\mu \dot{\mathbf{H}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\operatorname{div} \epsilon \mathbf{E} = 4\pi\rho \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$\operatorname{div} \mu \mathbf{H} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

\mathbf{E} is the strength of the electric field, \mathbf{H} the strength of the magnetic field, \mathbf{F} the strength of the impressed E.M.F., ρ the density of the electric charge, ϵ and μ the dielectric constant and the permeability; for the sake of simplicity we will assume that ϵ and μ are uniform in the whole space. From all the vector fields which appear in the equations only $\mathbf{F}(x, y, z, t)$ is given, the rest are unknown. The tangential components of \mathbf{E} and \mathbf{H} must be continuous on any surface of discontinuity such as the boundary between conductor and free space,

From (2) and (3) follows the energy relation

$$\frac{1}{8\pi} \frac{d}{dt} \int \mu H^2 dv + \frac{1}{8\pi} \frac{d}{dt} \int \epsilon E^2 dv + \int \sigma (E+F)^2 dv + \frac{c}{4\pi} \int [E, H]_n dS \\ = \int \sigma (E+F) F dv. \quad (6)$$

In (6) $\frac{\mu H^2}{8\pi}$ is the density of the magnetic energy, $\frac{\epsilon E^2}{8\pi}$ the density of the electric energy, $\sigma(E+F) = J_c$ is the conduction current, $\sigma(E+F)^2 = \frac{J_c^2}{\sigma}$ the rate of heat developed per unit of volume. $\int [E, H]_n dS$ the rate of energy flowing out from the region considered (if negative it means an inflow), and, finally, $\sigma(E+F)F = J_c F$ is the rate of work done by the impressed force per unit of volume.

If the system can be treated as quasi-stationary and quasi-static (low frequencies), and the integration in (6) be carried out over the whole space, $\int [E, H]_n dS = 0$. This follows easily from the fact that the electric field, which depends on quasi-static charges, drops at least as $1/r^2$, and the same applies to the magnetic field.

If no condensers are present and the network is formed by coils only $\epsilon \dot{E}$ is negligible in the whole space (for slow changes), and if no coils are present and the network consists of condensers only, the magnetic displacement current $\mu \dot{H}$ is negligible in the whole space and $\text{curl } E = 0$. If both coils and condensers are present, the electric field can be regarded as concentrated between the condenser plates only, where $\mu \dot{H} \doteq 0$, and the magnetic field in the coils only, where $\epsilon \dot{E} \doteq 0$ (lumped fields). In this case E is derivable from a scalar potential

$$E = -\text{grad } \phi_0 \quad (7)$$

$$\phi_0 = \frac{1}{\epsilon} \int \frac{\rho dv}{r} \quad (\text{non-retarded}), \quad (8)$$

and H from a vector potential

$$\mu H = \text{curl } A_0 \quad (9)$$

$$A_0 = \frac{\mu}{c} \int \frac{J_c dv}{r} \quad (\text{non-retarded}). \quad (10)$$

Actually μH can always be derived from a vector potential by virtue of (5), but (10) will apply only for quasi-stationary currents. In the general case the dependence between the current and the vector potential is different (see later). In (10) appears J_c and not $J_c + \frac{\epsilon \dot{E}}{4\pi} = J$, the full current, the reason being that as far as the magnetic "field" is concerned at low frequencies, the displacement current is negligible.

The density of the magnetic energy u_m can be written as

$$u_m = \frac{1}{8\pi} \mathbf{H} \text{ curl } \mathbf{A}_0.$$

Using the vector-analytical relation

$$\mathbf{H} \text{ curl } \mathbf{A}_0 = \mathbf{A}_0 \text{ curl } \mathbf{H} - \text{div} [\mathbf{H}, \mathbf{A}_0],$$

we get

$$u_m = \frac{1}{8\pi} \mathbf{A}_0 \text{ curl } \mathbf{H} - \frac{1}{8\pi} \text{div} [\mathbf{H}, \mathbf{A}_0].$$

From Gauss's theorem it follows that

$$\int u_m dv = \frac{1}{8\pi} \int \mathbf{A}_0 \text{ curl } \mathbf{H} dv - \frac{1}{8\pi} \int_S [\mathbf{H}, \mathbf{A}_0]_n dS.$$

If the integration is carried out over the whole space (surface $S \rightarrow \infty$) the surface integral vanishes in the quasi-stationary case, and we get

$$\begin{aligned} U_m &= \frac{1}{8\pi} \int \mathbf{A}_0 \text{ curl } \mathbf{H} dv = \frac{1}{2c} \mathbf{A}_0 \mathbf{J}_e dv \\ &= \frac{1}{2c} \int_{\text{volume of conductor}}^{\infty} \mathbf{A}_0 \mathbf{J}_e dx = \text{total magnetic energy.} \quad (11) \end{aligned}$$

(Again it will be noticed that for $c \text{ curl } \mathbf{H}$ we inserted $4\pi \mathbf{J}_e$ and not $4\pi \mathbf{J}_e + \epsilon \dot{\mathbf{E}}$.)

It is important to realize that (11) is valid for slow variations only and that the integration must be carried out over the whole conductor.

In a similar way we obtain for the electric energy

$$u_e = -\frac{1}{8\pi} (\text{grad } \phi, \epsilon \mathbf{E}).$$

Now, by Gauss's theorem,

$$\int \text{div} (\phi \epsilon \mathbf{E}) dv = \int_S \phi \epsilon E_n dS,$$

so that

$$\begin{aligned} \int \epsilon \mathbf{E} \text{ grad } \phi dv + \int \phi \text{div} (\epsilon \mathbf{E}) dv &= \int \phi \epsilon \mathbf{E} dS, \\ 4\pi \int \phi \rho dv &= - \int \epsilon \mathbf{E} \text{ grad } \phi dv + \int \phi \epsilon E_n dS. \end{aligned}$$

If the integration is carried out over the whole space, the surface integral vanishes and we get

$$U_e = \frac{1}{2} \int_{\infty} \phi \rho dv = \frac{1}{2} \int \phi \rho dv = \text{total electric energy.} \quad (12)$$

The energy relation now takes the form

$$\frac{d}{dt} \left(\frac{1}{2c} \int \mathbf{A}_0 \mathbf{J}_e dv \right) + \frac{d}{dt} \frac{1}{2c} \int \phi_0 \rho dv + \int \frac{\mathbf{J}_e^2}{\sigma} dv = \int \mathbf{J}_e \mathbf{F} dv. \quad (13)$$

(The integration being over the whole conductor.)

From this equation it is but a short step to the circuit equation. Inserting (8) and (10) in (13) we get

$$\frac{\mu}{c^2} \iint \frac{\dot{\mathbf{J}}_e \mathbf{J}'_e dv dv'}{r} + \frac{1}{\epsilon} \iint \frac{\dot{\rho} \rho' dv dv'}{r} + \int \frac{\mathbf{J}_e^2}{\sigma} dv = \int \mathbf{J}_e \mathbf{F} dv. \quad (14)$$

In the simplest case when the circuit consists of a resistance, an inductance and a capacitance in series, we can regard the current as having the same value along the whole wire and as being uniform in the cross-section. Let $i = |\mathbf{J}_c|q$, where q is the cross-section of the wire (not necessarily constant). i is the total conduction current. Then

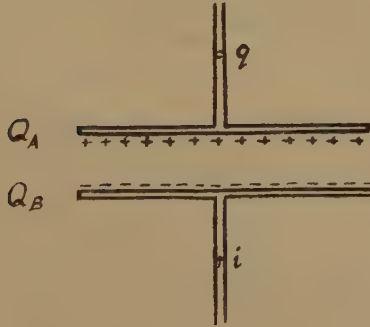
$$\iint \frac{\mathbf{J}_c \mathbf{J}_c' dv dv'}{r} = i i \iint \frac{dv dv' \mathbf{N} \mathbf{N}'}{r q^2}.$$

\mathbf{N} and \mathbf{N}' are unit vectors which indicate the direction in which the current flows in the points x, y, z and x', y', z' . $\mathbf{N} \mathbf{N}'$ is their scalar product. \mathbf{N} and \mathbf{N}' vary, of course, when we proceed along the wire, unless the wire is very thin and straight.

The charges will appear only over the plates of the condenser, where ρ appears as a surface charge ρ_s and its volume integral becomes a surface integral.

$$\text{Let} \quad Q_A = \int \rho_{sA} dS_A = -Q_B = \int \rho_{sB} dS_B = Q. \quad (\text{See fig. 1.})$$

Fig. 1.



Assuming the surface charge uniformly distributed over the plates of the condenser, we shall have

$$\iint \frac{\dot{\rho} \rho' dv dv'}{r} = 2Q\dot{Q} \iint \frac{dS_A dS'_A}{S^2 r} - 2Q\dot{Q} \iint \frac{dS_A dS_B}{S^2 r}$$

(S area of each plate).

$$\begin{aligned} \text{Finally,} \quad \int \frac{J_c^2 dv}{\sigma} &= i^2 \int \frac{dv}{q^2 \sigma} = i^2 \int \frac{dl}{q \sigma}, \\ \int \mathbf{J}_c \mathbf{F} dv &= i \int \frac{\mathbf{F}_l dv}{q} = i \int \mathbf{F}_l dl. \end{aligned}$$

(\mathbf{F}_l is assumed uniform in the cross-section of the wire and acting along it.)

We thus get

$$i i \frac{\mu}{c^2} \iint \frac{\mathbf{N} \mathbf{N}' dv dv'}{r q^2} + Q\dot{Q} \frac{1}{\epsilon} \iint \frac{2dS_A dS'_A - 2dS_A dS_B}{S^2 r} + i^2 \int \frac{dl}{\sigma q} = i \int \mathbf{F}_l dl. \quad (15)$$

Using $i = \dot{Q}$, and dividing the whole equation by this quantity, we get

$$i \iint \frac{NN' dv dv'}{rq^2} + \frac{1}{\epsilon} \int i dt \iint \frac{2dS_A dS'_A - 2dS_A dS_B}{S^2 r} + i \int \frac{dl}{q\sigma} = \int \mathbf{F}_i dl = \Phi. \quad (16)$$

From now on we can introduce complex quantities (they cannot be introduced in the energy equation as it is quadratic). Let Φ , the total E.M.F., be periodic and represented by the real part of $\Phi_a e^{j\omega t}$. Without loss of generality we may assume Φ_a real. For i we write $i_a e^{j\omega t}$. i_a will in general be complex, and it is the real part of $i_a e^{j\omega t}$ which matters. We then get

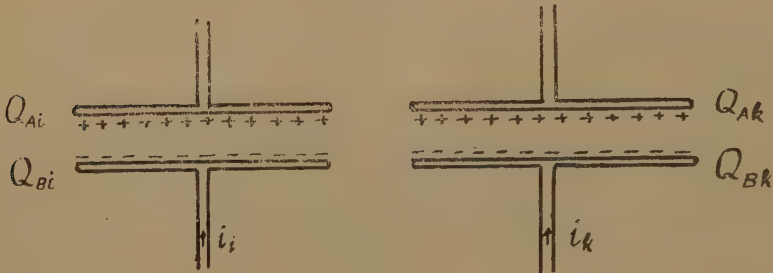
$$j\omega L i_a + \frac{i_a}{j\omega C} + i_a R = \Phi_a, \quad (17)$$

where

$$L = \frac{\mu}{c^2} \iint \frac{NN' dv dv'}{rq^2}, \quad C^{-1} = \frac{1}{\epsilon} \iint \frac{2dS_A dS'_A - 2dS_A dS_B}{S^2 r}, \quad R = \int \frac{dl}{q\sigma}. \quad (18)$$

The method outlined to obtain the circuit equation (17) for a single

Fig. 2.



mesh can easily be generalized for n meshes which may even have branches in common. The energy equation leads to

$$\sum_{i=1}^n \sum_{k=1}^n L_{ik} i_i i_k + \sum_{i=1}^n \sum_{k=1}^n C_{ik}^{-1} Q_i \dot{Q}_k + \sum_{i=1}^n \sum_{k=1}^n R_{ik} i_i i_k = \sum_{i=1}^n \Phi_i i_i$$

or

$$\sum_{k=1}^n L_{ik} i_k + \sum_{k=1}^n C_{ik}^{-1} \int i_k dt + \sum_{k=1}^n R_{ik} i_k = \Phi_i. \quad (19)$$

Hence

$$\sum_{k=1}^n Z_{ik} i_k = \Phi_i,$$

where

$$Z_{ik} = j\omega L_{ik} + j\omega C_{ik}^{-1} + R_{ik} \quad (20)$$

and

$$L_{ik} = \frac{\mu}{c^2} \iint \frac{(N_i N_k) dv_i dv_k}{rq_i q_k} \quad (21)$$

$$C_{ik}^{-1} = \frac{1}{\epsilon} \iint \frac{dS_{Ai} dS_{Ak}}{S_{Ai} S_{Ak} r} + \iint \frac{dS_{Bi} dS_{Bk}}{S_{Bi} S_{Bk} r} - \iint \frac{dS_{Ai} dS_{Bk}}{S_{Ai} S_{Bk} r} - \iint \frac{dS_{Bi} dS_{Ak}}{S_{Bi} S_{Ak} r} \quad (22)$$

$$R_{ik} = \frac{dl_{ik}}{q\sigma} \quad (23)$$

(dl_{ik} are elements which are common to the i th and k th mesh.)

III. THE ENERGY RELATION IN TERMS OF THE RETARDED VECTOR AND SCALAR POTENTIALS.

Before we derive the exact circuit theory for small loops and short pieces of wires and networks composed of such, it will be necessary to express the energy relation in terms of the retarded vector and scalar potentials. The use of these potentials which satisfy an inhomogeneous wave equation and not Poisson's equation, as is the case with the non-retarded potentials, is well known. The energy relation, however, is seldom expressed in terms of these potentials. At any rate no direct use seem to have been made of the equation to be derived, in antenna calculations. Unlike (13) the expression to be derived now will be exact.

$$\text{Again, let} \quad \mu \mathbf{H} = \text{curl } \mathbf{A}. \quad \dots \dots \dots (24)$$

$$\text{From (3) follows} \quad -c \text{ curl } \mathbf{E} = \text{curl } \dot{\mathbf{A}}.$$

$$\text{Thus} \quad \mathbf{E} = -\text{grad } \phi - \frac{\dot{\mathbf{A}}}{c}. \quad \dots \dots \dots (25)$$

ϕ is so far an entirely arbitrary scalar and so is $\text{div } \mathbf{A}$. Inserting (24) and (25) in (3) we get

$$c \text{ curl curl } \mathbf{A} + \epsilon \mu \left(\text{grad } \dot{\phi} + \frac{\ddot{\mathbf{A}}}{c} \right) = 4\pi \mu \mathbf{J}_c. \quad \dots \dots \dots (26)$$

From (4) and (25) we get

$$-\epsilon \text{ div } (\text{grad } \phi) - \epsilon \text{ div } \frac{\dot{\mathbf{A}}}{c} = 4\pi \rho. \quad \dots \dots \dots (27)$$

As we are free to dispose of $\text{div } \mathbf{A}$ and ϕ in any manner we choose, we use

$$\text{div } \mathbf{A} = -\frac{\epsilon \mu}{c} \dot{\phi}. \quad \dots \dots \dots (28)$$

This choice serves a double purpose: to eliminate ϕ from (26) and \mathbf{A} from (27). The result is

$$\text{curl curl } \mathbf{A} - \text{grad } (\text{div } \mathbf{A}) + \frac{\epsilon \mu}{c^2} \ddot{\mathbf{A}} = \frac{4\pi \mu}{c} \mathbf{J}_c. \quad \dots \dots \dots (29)$$

$$-\nabla^2 \phi + \frac{\epsilon \mu}{c^2} \ddot{\phi} = \frac{4\pi \rho}{\epsilon}. \quad \dots \dots \dots (30)$$

(29) is the vector wave equation for \mathbf{A} and (30) the scalar wave equation for ϕ . In cartesian co-ordinates, the vector wave equation splits into three scalar wave equations, but this is known to happen only in these co-ordinates.

As is well known and can be found in most text-books on Maxwell's theory, (29) and (30) lead to the following integral relations:

$$\mathbf{A}(t) = \frac{\mu}{c} \int \frac{\mathbf{J}(t^*) dv}{r}, \quad \dots \dots \dots (31)$$

$$\phi(t) = \frac{1}{\epsilon} \int \frac{\phi(t^*) dv}{r}, \quad \dots \dots \dots (32)$$

In (38) the surface which bounds the region must not cut through any of the conductors.

Inside a homogeneous conductor charges cannot accumulate, if E.M.F.s are not present. The equation of continuity $\text{div } \mathbf{J}_c + \dot{\rho} = 0$ gives

$$\text{div } \sigma \mathbf{E} + \text{div } \epsilon \dot{\mathbf{E}} = 0,$$

$$\frac{\sigma}{\epsilon} \text{div } \epsilon \mathbf{E} + \text{div } \epsilon \dot{\mathbf{E}} = 0,$$

$$\frac{\sigma}{\epsilon} \rho + \dot{\rho} = 0; \quad \rho = \rho_0 e^{-\frac{\sigma}{\epsilon} t}.$$

Thus independently of any oscillation (in which we shall be ultimately interested) the density of the charge decreases exponentially and becomes zero. Charges can accumulate only over the surface. When E.M.F.s are present charges will accumulate also inside the conductor wherever $\text{div } \mathbf{F} \neq 0$, because the equation of continuity gives

$$\text{div } \sigma(\mathbf{E} + \mathbf{F}) + \dot{\rho} = 0,$$

$$\frac{\sigma}{\epsilon} \rho + \dot{\rho} = -\sigma \text{div } \mathbf{F}.$$

If $\text{div } \mathbf{F}$ is periodic we shall have periodic charge fluctuations inside the conductor. We shall assume \mathbf{F} uniform inside the conductor, so that even with E.M.F., charge will appear in each single conductor only over the surface.

Thus (38) takes the form

$$\frac{1}{c} \int \dot{\mathbf{A}} \mathbf{J}_c dv + \int \dot{\rho}_s \phi dS + \int \frac{\mathbf{J}_c^2}{\sigma} dv = \int \mathbf{J}_c \mathbf{F} dv. \quad (39)$$

ρ_s is the charge which accumulates over the surface of the conductor, so that $\dot{\rho}_s$ is equal to \mathbf{J}_{cn} at the actual boundaries of the conductor. (n points in the direction perpendicular to the surface, outwards.)

Before we begin with the next paragraph, it may be added that

$$-\int \mathbf{E} \mathbf{J}_c dv + \int \frac{\mathbf{J}_c^2}{\sigma} dv = \int \mathbf{J}_c \mathbf{F} dv \quad (40)$$

(obtained in the course of deriving (36)) shows that the amount contributed at any moment by the E.M.F. to the field is given by

$-\int \mathbf{E} \mathbf{J}_c dv$. This relation has served a basis for another theory of antennæ, the so-called "counter E.M.F. method" which will be discussed briefly in § VIII. From (40) follows that the average radiated power is given by

$$-\frac{1}{2} \text{Re} \int \mathbf{E} \bar{\mathbf{J}}_c dv.$$

(Re=Real part. The bar denotes the conjugate complex.)

IV. THE EXACT CIRCUIT THEORY FOR SMALL LOOPS, SHORT AND THIN PIECES OF WIRE AND NETWORKS OF SUCH.

(a) Let us consider a very small loop. In such a loop we may assume the current as having the same value all along it, particularly as we assume F to be uniform inside the wire. Introducing (31) in (39), and noting that as the integration is carried out over the whole loop $J_{en} = \rho_s = 0$,

$$\text{we get} \quad \frac{\mu}{c^2} \iint \frac{J'_e(t) \dot{J}_e(t^*) dv dv'}{r} + \int \frac{J_e^2(t)}{\sigma} dv = \int J_e(t) F dv.$$

Again, assuming J_e as uniform in the cross-section and denoting the total current $|J_e|q$ by $i(t)$, we get

$$\frac{\mu i}{c} \iint \frac{\dot{J}_e(t^*) N dv dv'}{qr} + i^2 \int \frac{dl}{\sigma q} = i \int F_l dl.$$

After dividing by i and introducing complex quantities :

$$F_l = F_{la} e^{j\omega t}, \quad J_e(t) = J_{ea} e^{j\omega t}, \quad J_{ea}(t^* J_{ea}) e^{j\omega(t - r/c)}, \quad \dot{J}_e(t^*) = J_{ea} j\omega e^{j\omega(t - r/c)},$$

we obtain, after dropping the suffix a , as it is understood that the time is contained only in the common factor $e^{j\omega t}$:

$$\frac{j\omega\mu i}{c^2} \iint \frac{(NN') e^{-j\frac{\omega}{c}r} dv dv'}{q^2 r} + i \int \frac{dl}{\sigma q} = \int F_l dl.$$

Thus we again obtain an equation of the form

$$j\omega Li + Ri = \Phi,$$

where

$$L = \frac{\mu}{c^2} \iint \frac{e^{-j\frac{\omega}{c}r} NN' dv dv'}{rq^2}, \quad \dots \dots \dots (41)$$

$$R = \int \frac{dl}{\sigma q}. \quad \dots \dots \dots (42)$$

As will be seen the self-inductance is complex. Its imaginary part will give a power term in phase with the E.M.F. This term gives the radiated power.

To get a better idea of the complex self-inductance it will be instructive to follow L. Brillouin and expand the exponential in terms of $\frac{\omega r}{c} = \frac{2\pi}{\lambda} r$.

$$\frac{e^{-j\frac{\omega}{c}r}}{r} = \frac{1}{r} \left(1 - jr \frac{2\pi}{\lambda} - \frac{r^2}{2} \left(\frac{2\pi}{\lambda} \right)^2 + \frac{jr^3}{6} \left(\frac{2\pi}{\lambda} \right)^3 + \dots \right).$$

The constancy of current at any moment along the wire is at any rate justified only if $r \ll \lambda$. Therefore the series converges very quickly. The first term of L is

$$L_1 = \frac{\mu}{c^2} \iint \frac{NN' dv dv'}{rq^2}.$$

Comparing this expression with (18) it will be seen that it represents

the ordinary self-inductance. For a circular loop of a diameter d and a cross-sectional radius r_0

$$L_1 = \frac{2\pi d\mu}{c^2} \left(\log \frac{4d}{r_0} - \frac{7}{4} \right).$$

In calculating the subsequent terms we may assume the wire as infinitely thin, as the corresponding integrals do not diverge in connection with $r_0=0$.

$$L_2 = -\frac{j\mu}{c^2} \frac{2\pi}{\lambda} \iint dN dN' = 0$$

for a planar loop.

$$L_3 = -\frac{\mu}{2c^2} \left(\frac{2\pi}{\lambda} \right)^2 \iint r dN dN'.$$

The value of this integral for a circular loop is $\frac{\mu}{2c^3} \left(\frac{2\pi}{\lambda} \right)^2 \frac{16\pi d^3}{3}$. It is real and will represent a small addition to L_1 .

We obtain, further,

$$L_4 = \frac{j\mu}{6c^3} \left(\frac{2\pi}{\lambda} \right)^3 \iint r^2 dL dL' = -\frac{j\mu}{6c^2} \left(\frac{2\pi}{\lambda} \right)^3 4 (\text{area})^2.$$

Fig. 3.



Neglecting higher terms we thus get ($a=\text{area}$):

$$j\omega \left[L_1 + L_3 - j \frac{2a^2}{3c^2} \left(\frac{2\pi}{\lambda} \right)^3 \right] i + Ri = \Phi.$$

The radiated power will be

$$\frac{2}{3} \frac{(\text{area})^2}{c^2} \left(\frac{2\pi}{\lambda} \right)^3 \omega i_{\text{eff}}^2.$$

The radiation resistance is therefore to a first approximation $\frac{2}{3} \frac{\omega^4}{c^5} (\text{area})^2$. This is, of course, the well-known radiation resistance of a magnetic dipole, since $\text{area } i$ is the magnetic moment of the loop.

(b) Next let us consider a short and relatively thin piece of wire with uniform E.M.F. and let us assume the current flowing along the wire and hence charging its ends with oscillating charges. It will be convenient to

denote distances between points in different end-faces by r' and between points within an end-face by r'' . In (39)

$$\begin{aligned} \int \phi \dot{\rho}_S dS = & \iint \frac{\rho_{SA}^* \rho_{SA}' dS dS'}{r} = \iint \frac{\rho_{SA}^* \rho_{SA}' dS_A dS_A'}{r''} + \iint \frac{\rho_{SA}^* \rho_{SB} dS_A dS_B}{r'} \\ & + \iint \frac{\rho_{SB}^* \rho_{SB} dS_B dS_B'}{r''} + \iint \frac{\rho_{SB}^* \rho_{SA} dS_A dS_B}{r'}. \end{aligned}$$

Let $\rho_{SA} \cdot q_A = Q_A$, etc.

$$\dot{Q}_A = -\dot{Q}_B = |J_{cn'}| q_A = -|J_{cn''}| q_B = i.$$

Then

$$\begin{aligned} \int \phi \dot{\rho}_S dS = & \frac{\dot{Q}_A}{\epsilon} \iint \frac{\rho_{SA}^* dS_A dS_A'}{r'' q_A} + \frac{\dot{Q}_B}{\epsilon} \iint \frac{\rho_{SA}^* dS_A dS_B}{r' q_B} \\ & + \frac{\dot{Q}_B}{\epsilon} \iint \frac{\rho_{SB}^* dS_B dS_B'}{r'' q_B} + \frac{\dot{Q}_A}{\epsilon} \iint \frac{\rho_{SB}^* dS_A dS_B}{r' q_A}. \end{aligned}$$

After dividing by the current and introducing complex quantities, whereby

$$\rho_{SA}^* q_A = \rho_{SA} q_A e^{j\omega(t-r/c)} = Q_A e^{j\omega(t-r/c)} = \frac{i}{j\omega} e^{j\omega(t-r/c)}, \text{ etc.,}$$

we get

$$\begin{aligned} \frac{i}{j\omega} \left[\iint \frac{e^{-\frac{j\omega r''}{c}}}{r'' q_A^2} dS_A dS_A' - \iint \frac{e^{-\frac{j\omega r'}{c}}}{r' q_A q_B} dS_A dS_B \right. \\ \left. + \iint \frac{e^{-\frac{j\omega r''}{c}}}{r'' q_B^2} dS_B dS_B' - \iint \frac{e^{-\frac{j\omega r'}{c}}}{r' q_A q_B} dS_A dS_B \right] = \frac{i}{j\omega} C^{-1}. \quad (43) \end{aligned}$$

C is a complex capacitance.

As an example, let the current flow in a straight piece of wire of length l and diameter d and let $q_A = q_B = q$. Again we can expand the exponential in powers of r'/c and r''/c . The following integrals will appear:

$$+ \iint \frac{dS_A dS_A'}{r''} = f_1(d) q^2,$$

where $f_1(d) \doteq \frac{2}{d}$.

$$+ \iint \frac{dS_A dS_B}{r'} \doteq \frac{q^2}{l}.$$

$$\iint dS_A dS_B,$$

without r'' will cancel with $\iint dS_A dS_B$ without r' .

$$- \iint dS_A dS_A' r'' = q^2 f_2(d),$$

where $f_2(d) \doteq -\frac{d}{2}$.

$$-\iint dS_A dS_B r' \doteq -q^2 l.$$

$$j \iint dS_A dS_A r''^2 = j q^2 f_3(d),$$

$$\text{where } f_3(d) \doteq \left(\frac{d}{2}\right)^2.$$

$$j \iint dS_A dS_B r''^2 \doteq j q^2 l^2, \text{ etc.}$$

All the integrals are to be divided by q^2 , and the capacitive part gives

$$\frac{j}{j\omega} \left[\left(2f_1(d) - \frac{2}{l} \right) + \frac{\omega^2}{c^2} (l - f_2(d)) + \frac{2j\omega^3}{6c^3} (-l^2 + f_3(d)) + \dots \right].$$

As to the inductive part we shall have

$$L_1 = \frac{2l\mu}{c^2} \left(\log \frac{4l}{d} - \frac{3}{4} \right).$$

(The well-known ordinary self-inductance of a straight piece of wire.)

$$L_2 = -\frac{j\mu}{c^2} \frac{2\pi}{\lambda} \int_0^l \int_0^l dz dz' = -\frac{j\mu}{c^2} \frac{2\pi}{\lambda} l^2.$$

$$L_3 = \frac{\mu}{c^2} \int_0^l \int_0^l (z - z') dz dz' + \int_z^l \int_0^l (z' - z) dz dz' = 0.$$

$$L_4 = \frac{j\mu}{6c^3} \left(\frac{2\pi}{\lambda} \right)^3 \int_0^l \int_0^l (z - z')^2 dz dz' = \frac{j\mu}{6c^2} \left(\frac{2\pi}{\lambda} \right)^3 \frac{l^4}{6}, \text{ etc.}$$

$$\therefore j\omega L = j\omega\mu \left[\frac{2l}{c^2} \left(\log \frac{4l}{d} - \frac{3}{4} \right) - \frac{j}{c^2} \left(\frac{2\pi}{\lambda} \right) l^2 + \frac{j}{36c^2} \left(\frac{2\pi}{\lambda} \right)^3 l^4 + \dots \right].$$

Altogether the imaginary part which gives the power term connected with the radiation will be

$$\frac{2}{3} \left(\frac{2\pi}{\lambda} \right)^2 \frac{l^2}{c} + \frac{1}{3} \left(\frac{2\pi}{\lambda} \right)^2 \frac{f_3(d)}{c} - \frac{1}{36} \left(\frac{2\pi}{\lambda} \right)^4 \frac{l^4}{c} + \dots$$

The last two terms are small compared with the first, which is the well-known radiation resistance of a Hertzian dipole.

(c) Let us assume that a network consists of n current elements. Let two typical among them be $A_i B_i$ and $A_k A_k$. The distance between them will not be restricted in any way.

The mutual inductance between them will be

$$L_{ik} = \frac{\mu}{c^2} \iint \frac{N_i N_k e^{-j \frac{\omega}{c} r_{ik}} dv_i dv_k}{r_{ik} q_i q_k}.$$

In the limit for very small current elements we shall have

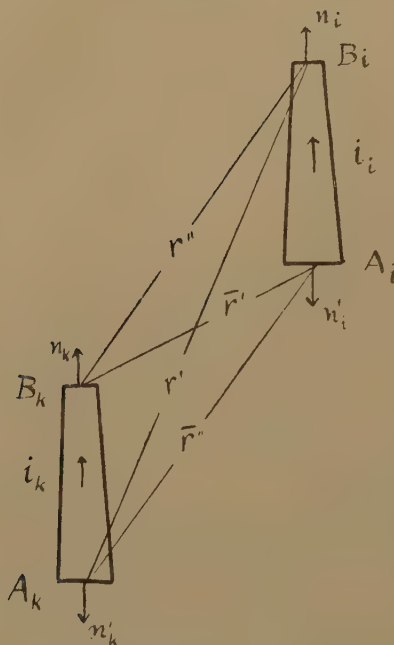
$$L_{ik} = \frac{\mu}{c^2} \frac{e^{-j \frac{\omega}{c} r(x, y, z; x', y', z')}}{r(x, y, z; x', y', z')} N_i N_k dl_i dl_k. \quad (44)$$

x, y, z is the position of the current element dl ; and x', y', z' of dl_k . $N_i = N(x, y, z)$ and $N_k = N(x', y', z')$.

The mutual capacitance is obtained as follows. From $\int \phi \dot{\rho}_S dS$, we get (see fig. 4) the mutual term

$$\begin{aligned} J_{ni} q_{Bi} \iint \frac{\rho_{SBk}^* dS_{Bi} dS_{Bk}}{r'' q_{Bi}} + J_{cn i} q_{Bi} \iint \frac{\rho_{SAk}^* dS_{Bi} dS_{Ak}}{r' q_{Bi}} \\ + J_{cn' i} q_{Ai} \iint \frac{\rho_{SBk}^* dS_{Ai} dS_{Bk}}{\bar{r}' q_{Ai}} + J_{cn' i} q_{Ai} \iint \frac{\rho_{SAk}^* dS_{Ai} dS_{Ak}}{\bar{r}'' q_{Ai}}. \end{aligned}$$

Fig. 4.



After dividing by $J_{cn i} q_{Bi}$, which is the fixed current element, and noticing that $J_{cn} q = -J_{cn'} q$, we get

$$\begin{aligned} \iint \frac{\rho_{SBk}^* dS_{Bi} dS_{Bk}}{r'' q_{Bi}} + \iint \frac{\rho_{SAk}^* dS_{Bi} dS_{Ak}}{r' q_{Bi}} \\ - \iint \frac{\rho_{SBk}^* dS_{Ai} dS_{Bk}}{\bar{r}' q_{Ai}} - \iint \frac{\rho_{SAk}^* dS_{Ai} dS_{Ak}}{\bar{r}'' q_{Ai}}. \end{aligned}$$

Introducing complex quantities and writing

$$\rho_{SBk} q_{Bk} = \frac{|J_{ck} q_{Bk}|}{j\omega} = \frac{i}{j\omega}, \quad \rho_{SAk} q_{Ak} = -\frac{|J_{ck}| q_{Ak}}{j\omega} = -\frac{i}{j\omega},$$

we get

$$\begin{aligned} & \frac{i}{j\omega\epsilon} \iint \frac{e^{-j\frac{\omega}{c}r''} dS_{B_i} dS_{B_k}}{r'' q_{B_i} q_{B_k}} - \frac{i}{j\omega\epsilon} \iint \frac{e^{-j\frac{\omega}{c}r'} dS_{B_i} dS_{A_k}}{r' q_{B_i} q_{A_k}} \\ & - \frac{i}{j\omega\epsilon} \iint \frac{e^{-j\frac{\omega}{c}\bar{r}''} dS_{A_i} dS_{B_k}}{\bar{r}'' q_{A_i} q_{B_k}} + \frac{i}{j\omega\epsilon} \iint \frac{e^{-j\frac{\omega}{c}\bar{r}'} dS_{A_i} dS_{A_k}}{\bar{r}' q_{A_i} q_{A_k}} = \frac{i}{j\omega\epsilon} C_{ik}^{-1}. \quad (45) \end{aligned}$$

As the current elements are very narrow the integration simply gives

$$\begin{aligned} & \frac{i}{j\omega\epsilon} \left[\frac{e^{-j\frac{\omega}{c}r''}}{r''} - \frac{e^{-j\frac{\omega}{c}r'}}{r'} - \frac{e^{-j\frac{\omega}{c}\bar{r}'}}{\bar{r}'} + \frac{e^{-j\frac{\omega}{c}\bar{r}''}}{\bar{r}''} \right] \\ & = -\frac{i}{j\omega\epsilon} \left[\left(\text{grad}_i \frac{e^{-j\frac{\omega}{c}r}}{r}, N_i \right)_{B_k} dl_i - \left(\text{grad}_i \frac{e^{-j\frac{\omega}{c}r}}{r}, N_i \right)_{A_k} dl_i \right]. \end{aligned}$$

The suffixes B_k and A_k indicate that the gradient is to be taken once from B_k as a centre and then from A_k as a centre. The difference can be expressed again as a gradient, this time along N_k . We get

$$\frac{i}{j\omega\epsilon} \left(\text{grad}_k \left(\text{grad}_i \frac{e^{-j\frac{\omega}{c}r}}{r}, N_i \right), N_k \right) dl_i dl_k.$$

Thus

$$C_{ik}^{-1} = \frac{1}{\epsilon} \left(\text{grad}' \left(\text{grad} \frac{e^{-j\frac{\omega}{c}r(x,y,z;x',y',z')}}{r(x,y,z;x',y',z')}, N(x,y,z) \right), N(x',y',z') \right) dl' dl'.$$

There are no mutual terms of resistance as the current elements have no current sub-elements in common. Only a self-resistance term $\frac{dl}{\sigma q}$ remains.

V. THE INTEGRAL EQUATION FOR THE CURRENT.

In this section we shall assume that as a result of the periodic E.M.F., the periodic current which flows in the conductors, flows along stationary lines. In general, in a three-dimensional conductor, the lines of flow will not be rigid, but will themselves be deformed periodically. We shall not consider this more general case.

If the conductor is subdivided into a great number n of current elements, we shall have

$$\begin{aligned} Z_{11} i_1 + Z_{12} i_2 + \dots + Z_{1n} i_n &= F_{11} dl_1 - \frac{i_1 dl_1}{q_1 \sigma} \\ Z_{21} i_1 + Z_{22} i_2 + \dots + Z_{2n} i_n &= F_{12} dl_2 - \frac{i_2 dl_2}{q_2 \sigma} \\ &\dots \dots \dots \\ Z_{n1} i_1 + Z_{n2} i_2 + \dots + Z_{nn} i_n &= F_{1n} dl_n - \frac{i_n dl_n}{q_n \sigma}. \end{aligned}$$

where

$$Z_{ik} = j\omega L_{ik} + \frac{1}{j\omega C_{ik}},$$

and L_{ik} and C_{ik} are given by (44) and (46).

The Z_{ik} are proportional to $dl_i dl_k$. Writing $Z_{ik} = K_{ik} dl_i dl_k$, we get

$$K_{k1} dl_k dl_1 |J_{c1}| q_1 + K_{k2} dl_k dl_2 |J_{c2}| q_2 + \dots + K_{kn} dl_k dl_n |J_{cn}| q_n \\ = \left(F_{ik} - \frac{|J_{ck}|}{\sigma} \right) dl_k \quad \dots \quad (47)$$

($k=1, 2, 3, \dots, n$.)

If the subdivision is carried out *ad infinitum*, the system of linear equation (47) becomes equivalent to an integral equation. After dividing (47) by dl_k and denoting $dl_i q_i$ by dv' we get

$$\int K(x, y, z; x', y', z') |J_c(x', y', z')| dv' = |F(x, y, z)| - \frac{J_c(x, y, z)}{\sigma}. \quad (48)$$

In (48)

$$K(x, y, z; x, y, z) = \frac{j\omega\mu e^{-j\frac{\omega}{c}r}}{c^2 r} NN' + \frac{1}{j\omega} \left(\text{grad}' \left(\text{grad} \frac{e^{-j\frac{\omega}{c}r}}{r}, N \right), N' \right). \quad (49)$$

$|F(x, y, z)| = F_N(x, y, z)$ is equal to the component of F in the direction of the current.

For $F=0$ we get an homogeneous integral equation the solution of which will give the free oscillations. In this case ω is complex and the solution will exist only for proper values of ω , the real part will give the frequency and the imaginary part the damping.

In the next section we shall deal with an example, the same which was treated by Hallén. It will be shown, for $\sigma=\infty$, that the integral equation (48) leads to the same integral equation as was used by him, so that we shall be able to use his solution. (The identity of the equations can be proved also for finite σ .)

VI. AN EXAMPLE: A STRAIGHT SYMMETRICAL ANTENNA OF NEGLIGIBLE OHMIC RESISTANCE.

Let the length of the antenna be $2l$ and its thickness $2r_0$. We will assume that the generator which enforces the oscillations is very small as compared with the antenna, so that the force F which represents the generator has a sharp peak at the middle of the antenna, and drops abruptly to both sides. The line integral of the force gives the total E.M.F. applied to the antenna (the "voltage"). We will assume it equal to $2v$. $\frac{2v}{i(0)}$ is the impedance of the antenna as viewed from the generator. In our calculation we assume the generator as impedanceless. Any impedance due to it can be included later.

If the antenna is long when compared with its width, so that $\frac{r_0}{l}$ is negligible and sets the upper limit of accuracy to be aimed in the calculation, we may regard N and N' as having the same direction parallel to

the z axis. It does not follow from this assumption that the current will turn out to be constant along the whole antenna. It simply means that we allow the charge to accumulate inside the wire instead of over its surface, as is actually the case through the slight bending of the lines of flow. The contribution of the charges to the field will be hardly effected by this assumption.

Although $\frac{r_0}{l}$ is to be neglected $\left(\log \frac{l}{r_0}\right)^{-1}$ is to be treated as a finite quantity, and not as zero.

Fig. 5.



As $NN'=1$, and N is the direction of the z axis,

$$(\text{grad } f(|R|), N) = \text{grad}_z f(|R|).$$

(Instead of the former r we will write $|R|$. r will be used in the following as a cylindrical co-ordinate.)

In cylindrical co-ordinates r, z, ϕ

$$|R| = \sqrt{(z-z')^2 + r^2 + r'^2 - 2rr' \cos(\phi - \phi')}.$$

The integral equation becomes

$$\int_{-l}^{+l} \int_0^{2\pi} \int_0^{r_0} \left[\frac{j\omega\mu}{c} \frac{e^{-j\frac{\omega}{c}|R|}}{|R|} + \frac{1}{j\omega\epsilon} \frac{\partial^2 e^{-j\frac{\omega}{c}|R|}}{\partial z \partial z' |R|} \right] J_0(z', r') r' d\phi' dz' + \frac{J_0(z, r)}{\sigma} = F_z(z, r). \quad (50)$$

By the cylindrical symmetry J_c does not depend on ϕ . The force F_z we will assume as depending both on z and r .

The integral equation is in two variables, r, z , while ϕ and ϕ' appear merely as parameters. In principle the whole equation can be integrated with respect to ϕ' as the kernel is a known function of this variable. In this way ϕ will also disappear. The integral with respect to ϕ' is, however, essentially elliptic, and we shall return to this question further down.

If σ is regarded as finite, there is no need to solve an integral equation of two variables. The dependence of J_c on r will hardly differ from that of an infinitely long wire. The skin effect in this case is well known and directly obtainable from Maxwell's equations when expressed in cylindrical co-ordinates. Thus $J_c(z', r')$ is a known function of r' and (50) could be integrated with respect to this variable. However, we shall not carry this calculation out, and assume $\sigma = \infty$, so that the current flows over the surface of the conductor. For high frequencies, which are alone of interest, this is to a high degree of approximation the case, and for a straight antenna the ohmic power loss is negligible when compared with the radiated power. Also the force F we will assume as acting over the surface only. Hence we have now the problem of a perfectly conducting thin walled tube.

In the following $J_c(z)$ will denote the total current per unit length over the surface so that $2\pi r_0 J_c(z) = i(z)$ will be the total conduction current at z .

The distance $|R|$ between two current elements becomes

$$|R| = \sqrt{(z-z')^2 + 4r_0^2 \sin^2 \frac{\phi-\phi'}{2}}. \quad (51)$$

Let us now consider the integration with respect to ϕ' . To start with, because of $\frac{\partial}{\partial z} = -\frac{\partial}{\partial z'}$, the integral equation can be written as

$$r_0 \int_{-l}^{+l} \int_0^{2\pi} \left[\frac{j\omega}{c} \frac{e^{-j\frac{\omega}{c}|R|}}{|R|} J_c(z') - \frac{1}{j\omega} \frac{\partial^2}{\partial z^2} \left(\frac{e^{-j\frac{\omega}{c}|R|}}{|R|} J_c(z') \right) \right] d\phi' dz' = F_z(z). \quad (52)$$

Therefore we must consider only the integral

$$\Phi(z) = -r_0 \int_{-l}^{+l} \int_0^{2\pi} \frac{e^{-j\frac{\omega}{c}|R|}}{|R|} J_c(z') d\phi' dz'. \quad (53)$$

This integral is convergent although the integrand is infinite at $z=z', \phi=\phi'$. The integral will diverge only for $r_0=0$. As the thickness of the wire has very little effect on the retardation, which is caused mainly through its length, we may write for

$$e^{-j\frac{\omega}{c}|R|} \rightarrow e^{-j\frac{\omega}{c}|z-z'|}.$$

It can also be shown that we may write approximately

$$\Phi(z) = -r_0 \int_{-l}^{+l} \int_0^{2\pi} \frac{e^{-j\frac{\omega}{c}|z-z'|} J_c(z') d\phi' dz'}{\sqrt{(z-z')^2 + 4r_0^2 \sin^2 \frac{\phi-\phi'}{2}}}$$

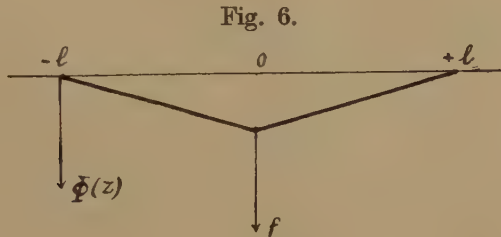
$$\div -2\pi r_0 \int_{-l}^{+l} e^{\frac{-j\omega}{c}|z-z'|} \frac{J_c(z') dz'}{\sqrt{(z-z')^2 + r_0^2}} = \int_{-l}^{+l} e^{\frac{-j\omega}{c}|z-z'|} \frac{i(z') dz'}{\sqrt{(z-z')^2 + r_0^2}} \quad (54)$$

(52) becomes
$$j\frac{\omega}{c}\Phi(z) + \frac{1}{j\omega} \frac{d^2}{dz^2} \Phi(z) = F_z(z),$$

or
$$\frac{d^2 \Phi(z)}{dz^2} + \frac{\omega^2}{c^2} \Phi(z) = j\omega F_z(z) = f(z). \quad (55)$$

Thus, the function $\Phi(z)$ satisfies a very simple differential equation, particularly in view of the shape of the force function that has been chosen. But to write the solution, we need the boundary condition, i. e., the value of $\Phi(z)$ for $z = \pm l$, and these are unknown.

The function $\Phi(z)$ is symmetrical ($\Phi(z) = \Phi(-z)$). The solution of (55) can be written as $\Phi(z) = \int G(z, z') f(z') dz'$, as is well known from the theory of Green's function. In our case the force $f(z)$ acts in the middle point $z=0$ only, such that the line integral of it is equal to $2j\omega v$ (see fig. 5).



The solution is therefore $\Phi(z) = G(0, z)$. The solution $\Phi(z)$ will have a discontinuous derivative

$$\left. \frac{d\Phi(z)}{dz} \right|_{z=+0} - \left. \frac{d\Phi(z)}{dz} \right|_{z=-0} = j\omega 2v. \quad (56)$$

(The discontinuity in the derivative of $\Phi(z)$ is best illustrated in the case of the simpler differential equation $\frac{d^2 \Phi(z)}{dz^2} = f(z)$, which is that of a simple string on which a force is acting in the middle point (see fig. 6). $\Phi(z)$ is to be interpreted as the displacement downwards.)

Thus it follows that a point generator which is represented by a force as described, introduces a discontinuity in the derivative of $\Phi(z)$. This is shown in fig. 7, where it is assumed for the sake of illustration that the function $\Phi(z)$ and the jump in the derivative are real. For $c = \infty$, the branches would consist of two straight lines. The term $\frac{\omega^2}{c^2} \Phi(z)$ in the differential equation bends the straight line into a curve which is composed from a sine and a cosine curve. This is readily seen from the

equation $\Phi''(z) + \frac{\omega^2}{c^2}\Phi(z) = 0$, which holds everywhere along the antenna except at $z=0$, where the discontinuity (56) is taking place. Thus

$$\left. \begin{aligned} \Phi(z) &= jc \left(a \cos \frac{\omega}{c} z + b \sin \frac{\omega}{c} z \right) & \text{for } z > 0 \\ \Phi(z) &= jc \left(a \cos \frac{\omega}{c} z - b \sin \frac{\omega}{c} z \right) & \text{for } z < 0 \end{aligned} \right\} \quad (57)$$

The solution is symmetrical and has a jump in the derivative which depends on b and not on a :

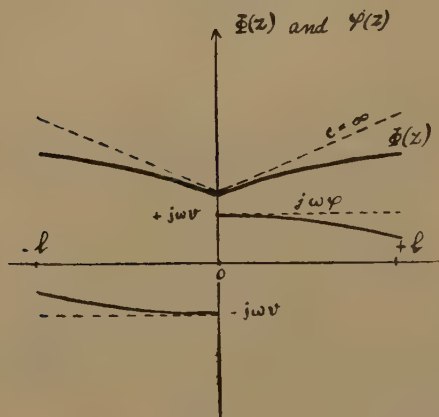
$$\Phi'_{+0}(z) - \Phi'_{-0}(z) = jc \left(b \frac{\omega}{c} + b \frac{\omega}{c} \right) = j\omega 2b.$$

Thus

$$b = v$$

$$(57) \text{ can be written } \Phi(z) = jc \left(a \cos \frac{\omega}{c} z + b \sin \frac{\omega}{c} |z| \right). \quad (58)$$

Fig. 7.



From now on the calculation can be taken over from Hallén's paper. Hallén does not introduce a force F , but introduces the generator by assuming a jump in the scalar potential equal to $2v$. As $j\omega\phi = \frac{d\phi(z)}{dz}$ this gives the same result as was obtained here by introducing F . The advantage of introducing F is that it can be introduced straight away in Maxwell's equations; this cannot be done with a jump in ϕ as ϕ does not appear in these equations. In fig. 7 $j\omega\phi$ is also indicated. For $c = \infty$; it will be constant along each half of the antenna.

As Hallén's paper is not easily accessible in this country, a very short account of his method to solve the integral equation will be included here.

Instead of dealing with the original integral equation (55) we deal with the problem expressed in (58), whereby the arbitrary constant a can be determined from the boundary condition $i(l) = i(-l) = 0$. (Had

we solved the original integral equation directly, we would encounter no arbitrary constant and the boundary condition would have been satisfied automatically.)

Hallén proceeds as follows. Put

$$\int_{-l}^{+l} \frac{i(z') e^{-j\frac{\omega}{c}|z-z'|}}{\sqrt{(z-z')^2 + r_0^2}} dz' = i(z) \int_{-l}^{+l} \frac{dz'}{\sqrt{(z-z')^2 + r_0^2}} + \int_{-l}^{+l} \frac{i(z') e^{-j\frac{\omega}{c}|z-z'|}}{\sqrt{(z-z')^2 + r_0^2}} - i(z) dz'. \quad (59)$$

In the last integral we may replace $\sqrt{(z-z')^2 + r_0^2}$ by $|z-z'|$ because the integral will remain convergent when dropping r_0 altogether. (For $z-z' < r_0$ r_0 will have but small influence, and for $z-z' \rightarrow 0$, the integrand will vanish.)

Now, we have the following identity

$$\int_{-l}^{+l} \frac{dz'}{\sqrt{(z-z')^2 + r_0^2}} = \log \frac{4l^2}{r_0^2} + \log \frac{l^2 - z^2}{l^2} + \log \frac{(l-z + \sqrt{(l-z)^2 + r_0^2})(l+z + \sqrt{(l+z)^2 + r_0^2})}{4(l-z)(l+z)}. \quad (60)$$

The expression under the last log is practically 1 for all $z \neq l$. At $z=l$ it increases, but here $i(z)$ is small and therefore we may neglect this log term altogether. Thus we get (writing $\Omega^{-1} = 2 \log \frac{2l}{r_0}$).

$$i(z) = -\frac{jc}{\Omega} \left(a \cos \frac{\omega}{c} z + v \sin \frac{\omega}{c} |z| \right) - \frac{1}{\Omega} \left[i(z) \log \frac{l^2 - z^2}{l^2} + \int_{-l}^{+l} \frac{i(z') e^{-j\frac{\omega}{c}|z-z'|}}{|z-z'|} - i(z) dz' \right]. \quad (61)$$

As will be seen the purpose of all this is to give the integral equation the shape of an equation of the second kind, rather than having to deal with an equation of the first kind. In this way an iteration method can be used to obtain the solution. (61) is, of course, not so exact as (55).

For $z=l$ (l is a positive quantity)

$$i(l) = 0 = -\frac{jc}{\Omega} \left(a \cos \frac{\omega}{c} l + v \sin \frac{\omega}{c} l \right) - \frac{1}{\Omega} \int_{-l}^{+l} \frac{i(z') e^{-j\frac{\omega}{c}(l-z')}}{(l-z')} dz'. \quad (62)$$

In order to make the subsequent iteration method "work" also for such lengths of antennæ where $\cos \frac{\omega}{c} l = 0$, and ensure $i(l) = 0$ in all approximations, we subtract (62) from (61):

$$i(z) = -\frac{jc}{\Omega} \left[a \left(\cos \frac{\omega}{c} z - \cos \frac{\omega}{c} l \right) + v \left(\sin \frac{\omega}{c} |z| - \sin \frac{\omega}{c} l \right) \right] - \frac{1}{\Omega} \left[i(z) \log \frac{l^2 - z^2}{l^2} + \int_{-l}^{+l} \frac{i(z') e^{-j\frac{\omega}{c}|z-z'|}}{|z-z'|} - i(z) dz' - \int_{-l}^{+l} \frac{i(z') e^{-j\frac{\omega}{c}(l-z')}}{l-z'} dz' \right]. \quad (63)$$

Now the iteration is as follows. As a first approximation we regard $i(z)$ as equal to the first square bracket (multiplied by $-\frac{jc}{\Omega}$). We then insert this expression for $i(z)$ or $i(z')$ in the right-hand side of (63) and carry out the integration. In this way we get an expression for $i(z)$ which consists of the first approximation; it is proportional to $\frac{1}{\Omega}$, plus an expression which is proportional to $\frac{1}{\Omega^2}$. Once more this new expression for $i(z)$ is inserted in the right-hand side of (63) and the integration carried out. A new term with $\frac{1}{\Omega^3}$ is thus added to the solution, and so forth. $\Omega = 2 \log \frac{2l}{r_0}$ is infinite for $r_0 = 0$. Therefore there is reason to believe that for thin enough wires the series will converge. For wires with $10^2 < \frac{2l}{r_0} < 10^5$, Ω varies between 10 and 35 approximately.

$$\text{Let} \quad \left. \begin{aligned} f_0(z) &= \cos \frac{\omega}{c} z & f_0(1) &= \cos \frac{\omega}{c} l \\ g_0(z) &= \sin \frac{\omega}{c} z & g_0(1) &= \sin \frac{\omega}{c} l \end{aligned} \right\}, \dots \dots \dots (64)$$

$$\text{and} \quad f_v(z) = -(f_{v-1}(z) - f_{v-1}(l)) \log \frac{l^2 - z^2}{l^2} \\ - \int_{-l}^{+l} \frac{f_{v-1}(z') - f_{v-1}(l) e^{-j \frac{\omega}{c} |z-z'|} - f_{v-1}(z) + f_{v-1}(l)}{|z-z'|} dz' \quad (65)$$

with a similar formula for the $g_v(z)$. We get

$$i(z) = -\frac{jc}{\Omega} a \left[f_0(z) - f_0(l) + \frac{f_1(z) - f_1(l)}{\Omega} + \frac{f_2(z) - f_2(l)}{\Omega^2} + \dots + \frac{f_v(z) - f_v(l)}{\Omega^v} \right] \\ - \frac{jc}{\Omega} v \left[g_0(z) - g_0(l) + \frac{g_1(z) - g_1(l)}{\Omega} + \frac{g_2(z) - g_2(l)}{\Omega^2} + \dots + \frac{g_v(z) - g_v(l)}{\Omega^v} \right]. \quad (66)$$

(66) satisfies (63) to a term

$$-\frac{jc}{\Omega^{v+2}} [a(f_{v+1}(z) - f_{v+1}(l) + v(g_{v+1}(z) - g_{v+1}(l)))].$$

The solution (66) still contains the undetermined constant a . By inserting (66) in (63) the constant a can be determined. Neglecting terms with $\frac{1}{\Omega^3}$ and higher powers we get

$$i(z) = \frac{jc}{\Omega f_0(l) + f_1(l) + \Omega^{-1} f_2(l)} v \left[g_0(l) f_0(z) - f_0(l) g_0(z) \right. \\ \left. + \frac{1}{\Omega} (g_0(l) f_1(z) - f_0(l) g_1(z) + g_1(l) f_0(z) - f_1(l) g_0(z)) \right]$$

$$+ \frac{1}{\Omega^2} (g_0(l)f_2(z) - f_0(l)g_2(z) + g_1(l)f_1(z) - f_1(l)g_1(z) + g_2(l)f_0(z) - f_2(l)g_0(z)) \Big]. \quad (67)$$

In particular, for the current at the generator follows

$$i(0) = \frac{jcv \left(\sin \frac{\omega}{c} l + \frac{\beta_1}{\Omega} + \frac{\beta_2}{\Omega^2} \right)}{\Omega \left(\cos \frac{\omega}{c} l + \frac{\alpha_1}{\Omega} + \frac{\alpha_2}{\Omega^2} \right)}. \quad (68)$$

$\alpha_1, \alpha_2, \beta_1, \beta_2$, are functions which result from the calculation as sketched above. They were calculated by Hallén. Recently R. King and F. G. Blake have recalculated them and plotted the results in excellent graphs.

The self-impedance of the antenna as it will appear to the generator is given by

$$Z_a = \frac{2v}{i(0)} = - \frac{j \frac{2\Omega}{c} \left(\cos \frac{\omega}{c} l + \frac{\alpha_1}{\Omega} + \frac{\alpha_2}{\Omega^2} \right)}{\sin \frac{\omega}{c} l + \frac{\beta_1}{\Omega} + \frac{\beta_2}{\Omega^2}}. \quad (69)$$

If we replace $\frac{2}{c}$ by 60, we get the impedance in ohms.

VII. RECEIVING ANTENNÆ.

Just as in a transmitting antenna where the generator which acts as a source of energy, is described by a force F , so also in a receiving antenna the receiving set which is joined to the antenna and acts as a sink of energy, can be represented by a force F . This force will oppose the current instead of encouraging it. It will be, essentially, in anti-phase with the current. Let the impedance of the receiving device be Z_r (as viewed from the terminals of the antenna). Then $\int F_r dl = 2v_r = -Z_r i(0)$.

The minus sign must be added because of the anti-phase relationship between the voltage across the terminals of the antenna and the current which flows through the receiver, should this latter absorb energy and not act as a generator. As a source of energy appears, of course, the incident electromagnetic wave. As the distance between transmitting antenna and receiving antenna is usually big any reaction of the latter on the former is negligible, and the introduction of an incident wave as a given function of space and time is fully warranted. It must be clearly understood that if the antenna were simply shorted, hardly any energy will be absorbed from the arriving wave. A small fraction of energy will be absorbed by the conductor and transformed into heat, but this will be negligible. The absorption of energy depends on the correct choice of the impedance Z_r .

Let it be clear that we cannot take into account the detailed construction of the receiver. We replace it just as we have done with the generator by a force F . In most cases F can be regarded as lumped

into a small space when it represents a point generator or a point receiver.

For a generator $\frac{\int F_1 dl}{i(0)} = \frac{2v}{i(0)} = Z_a$. For a receiver $\frac{\int F_1 dl}{i(0)} = -\frac{2v_r}{i(0)} = Z_r$.

In a transmitting antenna $2v$ is given and Z_a is to be calculated. In a receiving antenna Z_r is given and the voltage $2v_r$ is to be calculated. In both cases it amounts to calculating the current.

Let E_0, H_0 be the electromagnetic wave which arrives at the antenna. Through the presence of the antenna the field is modified by terms E', H' so that the total field is $E = E_0 + E', H = H_0 + H'$. Thus

$$\left. \begin{aligned} c \operatorname{curl} (H_0 + H') &= 4\pi\sigma(E_0 + E' + F) + \epsilon(\dot{E}_0 + \dot{E}') \\ c \operatorname{curl} (E_0 + E') &= -\mu(\dot{H}_0 + \dot{H}') \\ \operatorname{div} (E_0 + E') &= 4\pi\rho \\ \operatorname{div} (H_0 + H') &= 0 \end{aligned} \right\} \dots (70)$$

The conduction current is $J_c = \sigma(E_0 + E' + F)$. As the arriving wave itself satisfies Maxwell's equations, we have

$$\left. \begin{aligned} c \operatorname{curl} H_0 &= \epsilon \dot{E}_0 & \operatorname{div} \epsilon E_0 &= 0 \\ c \operatorname{curl} E_0 &= -\mu \dot{H}_0 & \operatorname{div} \mu H_0 &= 0 \end{aligned} \right\} \dots (71)$$

Subtracting (71) from (70), we get

$$\left. \begin{aligned} c \operatorname{curl} E' &= 4\pi\sigma(E_0 + E' + F) + \epsilon \dot{E}' & \operatorname{div} \mu H' &= 0 \\ c \operatorname{curl} H' &= -\mu \dot{H}' & \operatorname{div} \epsilon E' &= 4\pi\rho \end{aligned} \right\} \dots (72)$$

These equations are similar to those which we have examined before (2, 3, 4, 5), except for the appearance of $E_0 + F$ instead of F . Therefore, if we assume the lines of flow of the conduction current $\sigma(E_0 + E' + F)$ to be known we can use the integral equation

$$\frac{|J_c|}{\sigma} + \int K_r(x, y, z; x', y', z') |J_c(x', y', z')| dv' = |E_0 + F|. \quad (73)$$

In many cases it will be possible to assume the lines of flow as being practically the same both for the antenna when used for transmission as well as for reception, so that the kernel will be the same in both cases ($K_r = K$). In particular, this will be the case when the antenna consists of a thin wire.

From (73) follows quite generally that $J_c(x, y, z)$ must be linear in F and E_0 . In thin wires in particular, we may introduce the line integral of F which is v_r and the line integral of E_0 along the antenna $= v_i$ (incident voltage).

$$i(s) = f_1(s)v_r + f_2(s)v_i. \quad (74)$$

s is the distance from the generator along the wire while $f_1(s)$ and $f_2(s)$ are functions to be calculated from the integral equation. They depend only on the frequency, shape and length of the antenna and not on the receiver or the strength of the arriving wave. The meaning of $f_1(0)$ is quite simple. It is the self-admittance of the antenna (Z_a^{-1})

as it appears when the antenna is used for transmission. (For, if $E_0=0$, (73) is identical with our former integral equation.) Hence,

$$i(0) = \frac{v_r}{Z_a} + f_2(0)v_i. \quad (75)$$

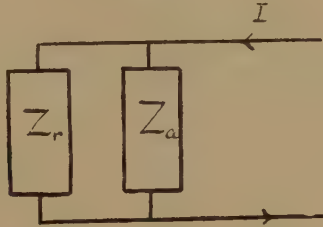
Now, for a receiving antenna $i(0) = -\frac{v_r}{Z_r}$, therefore

$$\frac{v_r}{Z_a} + \frac{v_r}{Z_r} = -f_2(0)v_i = I. \quad (76)$$

This equation has a very simple meaning: at a given frequency and a given strength of incident wave, an antenna acts as a constant current device (current= I). The current is supplied to Z_r and Z_a in parallel. Any power connected with Z_a means reradiation, and any power connected with Z_r real absorption in the receiver. Reradiation must be distinguished from reflection. See further down.

To absorb maximum power in the receiver we must have the resistive part of Z_r equal to the resistive part of Z_a (matching) and the reactive

Fig. 8.



parts equal in size but of opposite in sign (tuning). The maximum voltage across the terminals will appear when they are open circuited. For a given antenna, the only circuit element which we can vary is Z_r . The current supplied by the antenna is proportional to the incident voltage v_i , but it does not follow that it will increase with increasing length of antenna, because the function $f(s)$ is a complicated function of the length of the antenna.

The theory outlined here for receiving antennæ contains Hallén's theory for thin wires as a special case, in the same way as was shown for transmitting antennæ. Hallén has calculated the receiving properties of a straight symmetrical antenna, and a brief account of this calculation will be included here.

Let the component of the arriving wave along the wire be

$$E_{0z} = E_0 \cos \psi \cos \vartheta e^{-j\frac{\omega}{c}(\alpha + z \sin \vartheta)} \quad (77)$$

Then, if $q = \frac{\omega}{c} \sin \vartheta$ and $u = \frac{E_0 \cos \psi}{\frac{\omega}{c} \cos \vartheta}$, (78)

$$E_{0z} = \frac{1}{2}(E_{0z}(z) + E_{0z}(-z)) = u \frac{\omega}{c} \cos^2 \vartheta \cos qz \quad (79)$$

In (79) we have the expression for the symmetrical component of the arriving wave. This component alone is responsible for the symmetrical current through the antenna. An anti-symmetrical current will have a current node in the middle of the antenna, and no voltage will appear across a finite impedance connected to the terminals of the antenna.

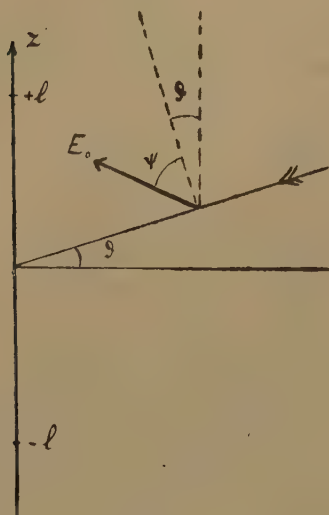
Instead of (55) the integral equation gives now

$$\frac{d^2\Phi(z)}{dz^2} + \frac{\omega^2}{c^2}\Phi(z) = j\omega F_z(z) + j\omega u \frac{\omega^2}{c^2} \cos^2 \vartheta \cos qz. \quad (80)$$

The solution is

$$\Phi(z) = j\omega \left(a \cos \frac{\omega}{c} z + v_r \sin \frac{\omega}{c} |z| + u \cos qz \right). \quad (81)$$

Fig. 9.



Following the same method as was described before, Hallén obtains (apart from the functions f and g appears a set of functions h , with $h_0 = \cos qz$ and subsequently apart from the functions α and β appear functions γ):

$$i(0) = \frac{\frac{jc}{\Omega} v_r \left(\sin \frac{\omega}{c} l + \frac{\beta_1}{\Omega} + \frac{\beta_2}{\Omega^2} \right) + u \left(\cos ql - \cos \frac{\omega}{c} l + \frac{\gamma_1}{\Omega} + \frac{\gamma_2}{\Omega^2} \right)}{\cos \frac{\omega}{c} l + \frac{\alpha_1}{\Omega} + \frac{\alpha_2}{\Omega^2}}. \quad (82)$$

Using (69) and $-\frac{2v_r}{i(0)} = Z_r$, (82) gives

$$\frac{2v_r}{Z_r} + \frac{2v_r}{Z_a} = - \frac{jcu \left(\cos ql - \cos \frac{\omega}{c} l + \frac{\gamma_1}{\Omega} + \frac{\gamma_2}{\Omega^2} \right)}{\cos \frac{\omega}{c} l + \frac{\alpha_1}{\Omega} + \frac{\alpha_2}{\Omega^2}} = I. \quad (83)$$

The maximum power absorbed for different lengths of antenna and for different angles of incidence ϑ was plotted by Hallén and may be found in his paper.

Before we proceed with the next paragraph a remark on what was called reradiation will be added. From (74) it will be seen that the current along a receiving antenna can be decomposed into two parts, $f_1(s)v_r$ and $f_2(s)v_i$. The total current is totally different in its distribution along the antenna from the current which would have resulted had the antenna been used for transmission, with a point generator in place of the receiver *. Only the component $f_1(s)v_r$, which is due to the presence of the receiver, has the same distribution as that of a transmission current, because it is due to F and not to E in its mathematical structure. On the other hand, the structure of the component $f_2(s)v_i$ depends only on the arriving wave and is independent of the receiver. In a shorted antenna this component alone will exist and give rise to a certain reflected field, with a certain reflected power. When the receiver is added the current is modified by a "transmission current" $f_1(s)v_r$ which gives a reradiated field. Had this component existed alone, the total reflected power would have been equal to the reradiated power. Only this reradiated power matters for matching. The actual reflected power which seems to come from the antenna, *i. e.* which depends on the *total* current which flows in the antenna, is obtained from the field E' , H' , which is the vectorial sum of the reflected field of the shorted antenna and the re-radiated field. Finally, it may be added that (75) may be written as

$$i(0)\left(1 + \frac{Z_r}{Z_a}\right) = f_2(0)v_i. \quad . \quad . \quad . \quad . \quad . \quad (84)$$

This form expresses it more clearly that the current depends only on the incident voltage.

VIII. REMARK ON THE COUNTER E.M.F. METHOD AND CARTER'S CIRCUIT RELATIONS.

In a paper entitled "On the Origin of the Radiation Resistance," L. Brillouin pointed out that it is not necessary to calculate the radiation by integrating Poynting's vector at infinity; it is possible to obtain this quantity by calculating the electric force in the conductor. This method is simpler and appeals more to the engineer who prefers to calculate all the quantities from the conditions at the generator and the attached conductors. In both methods the current must be known in its distribution and strength.

The electric force appears in the region of the E.M.F. as a counter force (E opposes F) and the work done by the E.M.F. is against this force. An elementary example is provided when a d.c. circuit is analysed, consisting of a loop of wire in part of which an E.M.F. is

* This does not affect the kernel, which depends only on the direction of the lines of flow. In both cases these are the same, namely along the wire.

impressed to represent the battery. In the region of the impressed force, E opposes F , while in the rest of the wire E and F have the same sense. No discontinuity in the sense of the magnetic field is present. Therefore Poynting's flow of energy is outwards in the region of the E.M.F. and inwards at the rest of the wire. A simple calculation shows in this case that the whole energy which flows out from the region of the E.M.F. is equal to the energy which flows into the rest of the wire to be transformed into heat. No radiation loss exists in this d.c. case.

The basis of Brillouin's idea is the formula (40) :

$$-\int E J_e dv + \int \frac{J_e^2}{\sigma} dv = \int J_e F dv.$$

(This relation holds for any region of integration.)

Brillouin neglects the ohmic resistance ($\sigma = \infty$) so that the work done by F is equal to $-\int E J_e dv$, and this represents the radiation.

$$\text{From} \quad -\int E J_e dv = \int J_e F dv \quad \dots \dots \dots (85)$$

it will be seen that for $\sigma = \infty$ an electric force inside the conductor will exist only in the region of F , where it will be equal to it in magnitude and opposite in direction. Outside the conductor E will be normal to it everywhere except on the boundaries of the region of F . At any rate, if E can be determined at the conductor the radiation can be found from $-\int J_e E dv$.

For thin wires we shall have

$$-\int i(s) E_s(s) ds = \int F_s(s) i(s) ds = W. \quad \dots \dots \dots (86)$$

Let us now assume that n wires are present. Let $E_m^{(k)}$ be the field along the k th conductor as contributed by the current and charges of the m th conductor. Then

$$-\sum_{k=1}^n \sum_{m=1}^n \int i_k(s) E_m^{(k)}(s) ds_k = W.$$

$$\text{Let} \quad i_k(s) = \text{Re}(I_k f_k(s) e^{j\omega t}) = \frac{I_k f_k(s) e^{j\omega t} + \overline{I_k f_k(s)} e^{-j\omega t}}{2}. \quad \dots \dots (87)$$

I_k is a complex constant while $f_k(s)$ is a complex function of s which we assume to have a unit modulus at a given point on each wire. $f_k(0) = 1$. This is the most general assumption we can make with regard to the forced oscillatory nature of the current. The field along the k th conductor which will result from $i_m(s)$ and the charges connected with it, will be

$$E_m^{(k)}(s) = \frac{1}{2} (I_m e_m(s) e^{j\omega t} + \overline{I_m e_m(s)} e^{-j\omega t}). \quad \dots \dots (88)$$

$e_m^{(k)}(s)$ is a complex function of s . The radiated power will be

$$W = -\frac{1}{4} \sum_{k=1}^n \sum_{m=1}^n \int [I_k f_k(s) \overline{I_m e_m^{(k)}(s)} + I_m e_m^{(k)}(s) \overline{I_k f_k(s)} + I_k f_k(s) I_m e_m^{(k)}(s) e^{2j\omega t} + \overline{I_k f_k(s)} \overline{I_m e_m^{(k)}(s)} e^{-2j\omega t}] ds_k. \quad \dots (89)$$

These equations are identical with Carter's circuit relations, as can easily be proved. From the significance of the kernel

$$K(s_k, s_m) = \frac{j\omega e^{-j\frac{\omega}{c}r_{km}}}{c^2 r_{km}} N_k N_m + \frac{1}{j\omega} \frac{\partial}{\partial s_m} \left(\frac{\partial}{\partial s_k} \frac{e^{-j\frac{\omega}{c}r_{km}}}{r_{km}} \right). \quad (99)$$

follows, after using $E_k = -\text{grad}_{s_k} \phi - \frac{\dot{A}_{s_k}}{c},$

$$-E_k^{(m)}(s_k) = \int_m K(s_k, s_m) i(s_m) ds_m. \quad (100)$$

Inserting (100) in (98) we get (93) so that the identity of (96) with (92) is proved.

Clearly, the circuit equations (92) cannot be used if the Z_{km} are not known. On the other hand, their determination involves the solution of the system of integral equations (96); once this is achieved, the currents are determined, and the circuit equations appear superfluous. In spite of this, they render great service, as often some reasonably approximate assumptions can be made about the current distribution $f_k(s_k)$. In his paper, Carter assumes that for $\frac{\lambda}{2}$ wires the current is sinusoidally distributed, so that the functions $f_k(s_k)$ are real: $\cos \frac{2\pi}{\lambda} z$, $z=0$ being the middle of the antenna. From Hallén's theory follows that this distribution will really take place for an infinitely thin wire, $\frac{\lambda}{2}, \frac{3}{2}\lambda, \frac{5}{2}\lambda, \dots$ long. (See equation 67.) For a $\frac{\lambda}{2}$ aerial Carter obtains the impedance $73.14 + j43$, and exactly this value also follows from Hallén's theory. But for any other length the current will not form a purely standing wave, and $i(s_k)$ will not be real. It is complex with the imaginary part being linearly independent of the real part. Thus, $\text{Re}(i(s)e^{j\omega t})$ will consist of two components in time quadrature, which have different distributions along the wire. For a straight antenna these two components were first calculated by L. V. King. He noted that it is impossible to satisfy the boundary condition, viz., that the parallel component of the electric force over the surface of the perfect wire should vanish, with one single current component (standing wave). For a $\frac{\lambda}{2}$ aerial with a sinusoidal standing current wave we actually get, even for a perfect wire, a longitudinal component of E along the wire, which seems to speak against Hallén's theory, or the validity of Carter's result. But this difficulty is only apparent. An infinitely thin wire is a limit to be approached, but never to be reached without some special analysis. Let us consider a very thin wire of finite thickness ($\sigma = \infty$). It will be found that at the surface the parallel component is zero, but that at a very small distance from it E will have a component parallel to the wire, bigger than a certain small value. Now if the wire is made thinner and thinner, the above distance gets smaller and smaller, while at the same time the parallel component

becomes relatively stronger. In the limit, when the wire is infinitely thin, the parallel component will reach the surface of the wire with some strength. This must not be regarded as contradicting the boundary condition.

References.

- (1) L. V. King, 'Philosophical Transactions of the Royal Soc.' 1936, vol. cccxxvi.
- (2) E. Hallén, *Nova Acta*, Upsala, 1938.
- (3) L. Brillouin, *Radioélectricité* (April 1922).
- (4) R. King and F. G. Blake, *Proc. I.R.E.* (July 1942).
- (5) P. S. Carter, *Proc. I. R. E.* vol. xx. (1932).

LII. Further Properties of Recurrent Exponential and Probability Function Pulse Wave-forms.

By F. F. ROBERTS, B.Sc. (Eng.) and J. C. SIMMONDS, Ph.D.*

[Received January 10, 1944.]

Summary.

As a sequel to the previous paper ⁽¹⁾† in which the recurrent exponential pulse was defined and shown to be of practically identical form to a repeated probability-function pulse, further examples are now given to demonstrate how these pulses may readily be manipulated mathematically whilst undergoing a variety of operations of practical interest.

The addition and multiplication of pulses are considered together with cross-talk problems associated with time division multichannel communication systems and possible methods of modulating pulses. The response of a network to a single pulse is dealt with briefly.

1. The Single Probability Function Pulse.

IN the varied applications of pulse signals, both for the transmission of information and for actually obtaining it from the pulses themselves (as in ionosphere sounding and its recent developments), the chief interest shifts back and forth between the wave-form of the pulse and its corresponding frequency spectrum as the viewpoint of design or application is altered. Any method for rapidly and accurately correlating the two views must, therefore, be of interest. It is the object of the present paper to describe and illustrate such a method, with particular reference to pulses of the "recurrent exponential" and "probability function" wave-forms defined and discussed in the previous contribution ⁽¹⁾.

* Communicated by the Authors.

† For numbered references see Bibliography.

Some general operations with the probability-function pulse will first be given, with examples of their use in computing cross-talk in a time division multiplex system, and a theorem on the response of a general type of network to such a pulse will be stated. The recurrent exponential pulse will then be considered, more particularly in connection with the possible types of modulation.

1.1. *Addition of Pulses.*

The pulse represented by

$$A_0 e^{-(t/T)^2}$$

has its peak value A_0 at the time origin. According to equation (15) of the previous paper ⁽¹⁾, the relative amplitude/frequency spectrum of this pulse may be written

$$A_0 T e^{-(\pi T f)^2}.$$

(The factor $\sqrt{\pi}$ will be omitted unless essential to the argument.)

A second pulse of peak value A_1 , following the first after an interval $t=t_1$, but otherwise having the same form, *i. e.* the same width factor T , is given by

$$A_1 e^{-(t-t_1)^2/T^2},$$

and in this case the Fourier integral for the spectrum gives

$$A_1 T e^{-j2\pi t_1 f} \cdot e^{-(\pi T f)^2}.$$

The resultant spectrum of $(n+1)$ such pulses is then given by the sum of the components, thus

$$T e^{-(\pi T f)^2} \cdot \sum_{r=0}^n A_r e^{-j2\pi t_r f},$$

remembering that $t_0=0$. In many cases only the magnitude of the relative spectrum is of interest; it is given by

$$|g(f)| = T e^{-(\pi T f)^2} \sqrt{\left(\sum_{r=0}^n A_r \cos 2\pi t_r f \right)^2 + \left(\sum_{r=0}^n A_r \sin 2\pi t_r f \right)^2},$$

which may be rewritten

$$|g(f)| = T e^{-(\pi T f)^2} \sqrt{\sum_{r=0}^n A_r^2 + 2 \sum_{\substack{r=0 \\ r \neq s}}^n A_r A_s \cos 2\pi(t_r - t_s)f}. \quad \dots (1)$$

The spectrum of the sum of any number of pulses of the same form but of arbitrary relative amplitudes and time delays is thus given by multiplying the spectrum of a single pulse by a simple function of the amplitudes and delays. For example, a single small pulse following a single large one produces a "ripple" on the spectrum, the period of the ripple being simply related to the delay between the two pulses. Another case of interest occurs when some of the pulses have the same amplitude, when it is evident from equation (1) that sharp dips may occur in the spectrum, in fact, if there are altogether only two equal pulses the spectrum dips to zero periodically.

1.2. Multiplication by an Independent Variable.

The simplest case is that where the variable is represented by a simple cisoidal function of time, and this again may be subdivided into two cases, depending on whether the wave-period is large or small compared with the parameter T of the pulse. Theoretically the two cases together are the inverse of that treated above; for the pulse multiplied by the cisoidal function may be written

$$e^{j2\pi f_0 t} \cdot e^{-(t/T)^2}$$

and the corresponding spectrum becomes

$$\frac{1}{2} T e^{-\pi^2 T^2 (f - f_0)^2}, \quad \dots \dots \dots (2)$$

i. e. the whole original spectrum is simply shifted along the frequency axis by an amount equal to the frequency of the multiplying function. When f_0 is small compared with $1/T$ the positive and negative frequency components are still substantially equal in magnitude. When f_0 is very large, on the other hand, the negative frequency components can be ignored and the spectrum becomes that normally associated with a pulse of a high frequency carrier.

1.3. Multiplication of One Pulse by Another.

Consider the product of two pulses of amplitudes A_1 and A_2 of the same width factor T but having a relative time difference t_0 . Clearly this may be written

$$A_1 e^{-(t/T)^2} \cdot A_2 e^{-(t-t_0)^2/T^2} = A_1 A_2 e^{-\frac{t_0^2}{2T^2} - 2(t - \frac{t_0}{2})^2/T^2} \dots \dots (3)$$

Now the last term in this expression has its peak value when $t = \frac{1}{2}t_0$ and has the same form as the original pulse. The product of the two pulses is thus a third pulse of the same general shape, but $\sqrt{2}$ times as sharp, located halfway on the time scale, and having its magnitude reduced by the factor $e^{-t_0^2/2T^2}$. The product is therefore practically unattenuated when t_0 is appreciably less than T , but falls off rapidly in magnitude as t_0 exceeds T . As will be explained below, this property of the pulse product finds application in the demodulation of multiple pulse signalling systems.

1.4. Time-Division Multiplex.

Multichannel communication systems operating on a basis of time division instead of the more usual frequency division have been described by a number of writers ^{(2), (3)}. One of the problems associated with these systems is that of cross-talk between the different channels. Now it is clearly quite feasible to produce such a system by employing exponential recurrent pulses of the type discussed here and in the previous paper ⁽¹⁾: —thus a number of pulses equal to the number of channels desired (with an extra pulse, perhaps, for synchronizing the receiving equipment) would be transmitted in regular sequence, each modulated in amplitude by the corresponding channel signal.

The first type of cross-talk to occur will then be that due to adjacent pulses being too close together compared with their widths. Assuming, then, the same pulse amplitudes and depths of modulation, the ratio of desired to undesired signal is seen, from the previous section, to be

$e^{t_0^2/2T^2},$
that is $t_0^2/2T^2$ nepers. (4)

This result shows that ample cross-talk attenuation from this cause is attainable with quite close pulses; *e. g.* about 40 decibels for $t_0=3T$.

However, another source of cross-talk arises in practice due to the finite band-width of the channel connecting transmitter and receiver. This band-width may be restricted at either the high frequency or low frequency ends or at both ends. These two types of cut-off result in quite different types of cross-talk in a multiplex system.

Taking the case of high frequency cut-off first, it will be remembered that the frequency spectrum of the probability-function pulse

$Ae^{-(t/T)^2}$ (5)

is another probability function

$ATe^{-(\pi Tf)^2}$ (6)

Now suppose that such a pulse is fed into a network whose amplitude/frequency spectrum can also be represented by this type of law, *e. g.*,

$e^{-(\pi Kf)^2}$, (7)

where K is a constant for the network. The output frequency spectrum from the network will then be

$ATe^{-\pi^2(K^2+T^2)f^2}$, (8)

which is again of the same form in f . It is seen, therefore, that the output wave-form is given by the pulse

$\frac{AT}{\sqrt{K^2+T^2}} e^{-t^2/(K^2+T^2)}$, (9)

that is, a pulse broadened and attenuated by the factor $\sqrt{1+(K/T)^2}$. From equation (4), therefore, such a high frequency cut-off characteristic will cause the cross-talk between adjacent pulses to be degraded by

$t_0^2/(K^2+T^2)$, (10)

assuming both pulses have the shape given by equation (9).

Turning now to low frequency cut-off, an almost equally simple analysis is possible, as follows. Consider a network whose amplitude/frequency characteristic may be represented by

$\{1-e^{-(\pi K_1 f)^2}\}$ (11)

and suppose the pulse, equation (5), is again applied to its input terminals. The output wave is evidently given by the sum of an unaltered pulse and an inverted, broadened and attenuated version of it corresponding to a network response

$e^{-(\pi K_1 f)^2}$,

From equations (5) and (9), therefore, the required response is

$$A \left\{ e^{-(t/T)^2} - \frac{T e^{-t^2/(K_1^2 + T^2)}}{\sqrt{K_1^2 + T^2}} \right\} (12)$$

Such a pulse is illustrated in fig. 1, where the value $K_1=3T$ has been chosen to make the effect obvious. It follows from equation (12) that

Fig. 1 (a).

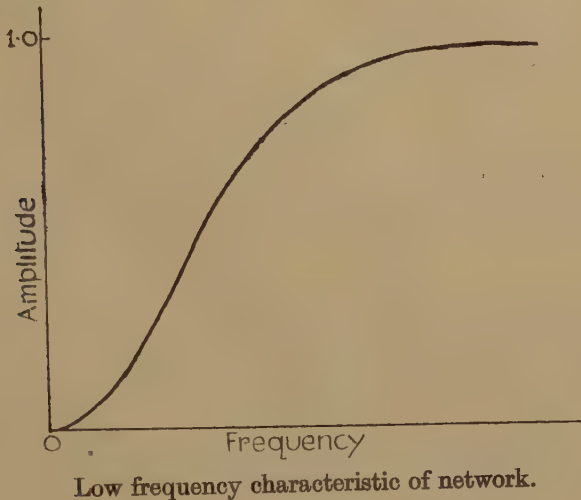
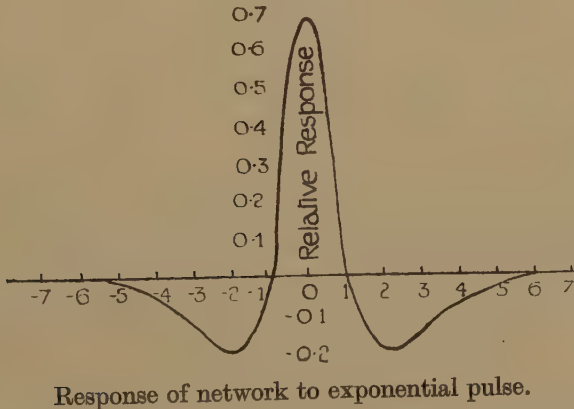


Fig. 1 (b).



any modulation impressed on the original pulse will also appear on the broadened, attenuated version at the output of the high-pass network. In practical cases, in order to make this cross-talk sufficiently small, the parameter K_1 must be made many times T , and this automatically makes the inverted response component so broad as to affect all channels of a multiplex system almost equally. The cross-talk ratio is thus given

by the relative magnitude of the inverted component, *i. e.* the cross-talk attenuation due to low frequency cut-off is

$$\log_e \sqrt{1 + (K/T)^2} \text{ nepers.} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

1.5. *Further Considerations.*

It will have been noted in the last section that the two types of cross-talk caused by high frequency and low frequency cut-off in the transmission link are in opposite phase. It is, therefore, possible to adjust the cut-off frequencies in order to benefit by the compensation of cross-talk that could be achieved. But it is clear that a simple combination of high frequency and low frequency cut-off laws of the types of equations (7) and (11) will permit compensation for only one particular pulse separation at a time, *i. e.* cross-talk to other channels would remain poor. However, the principle invoked in the above analysis of the low frequency cut-off effect may be extended to a more general form. For example, a network may have an amplitude/frequency response of the form

$$A_1 e^{-(\pi K_1 f)^2} + A_2 e^{-(\pi K_2 f)^2} + A_3 e^{-(\pi K_3 f)^2} \quad . \quad . \quad . \quad . \quad (14)$$

This corresponds to the previous case if $A_1=1$, $K_1=0$, $A_2=-1$ and $A_3=0$. The response of such a network to the pulse of equation (5) is given by

$$\frac{A_1 e^{-t^2/(K_1^2 + T^2)}}{\sqrt{1 + (K_1/T)^2}} + \frac{A_2 e^{-t^2/(K_2^2 + T^2)}}{\sqrt{1 + (K_2/T)^2}} + \frac{A_3 e^{-t^2/(K_3^2 + T^2)}}{\sqrt{1 + (K_3/T)^2}} \quad . \quad . \quad (15)$$

Thus, by suitably “shaping” the amplitude/frequency law of the network, it is possible to shape the output pulse until a more desirable combination of cross-talk and effective frequency band-width is obtained. Other authors⁽²⁾ have found that sharp high frequency cut-off above a certain point may, for instance, result in a net improvement in cross-talk compared with no effective cut-off at all until a much higher frequency ; this conclusion is confirmed in the following section, for the exponential recurrent pulse.

It might, finally, be pointed out that equations (14) and (15) might be even more generalized if corresponding to each frequency cut-off parameter K_n there is a phase delay t_n . The output wave-form in such a case may be written

$$\sum \frac{A_n e^{-(t-t_n)^2/(K_n^2 + T^2)}}{\sqrt{1 + (K_n/T)^2}} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

This extended analysis provides the basis for a new approach to the above and similar problems.

2. *The Modulated Exponential Recurrent Pulse.*

The preceding five sections have dealt particularly with pulses of the probability-function type, in order to show some of the analytical advantages of that formulation. As shown in the previous paper⁽¹⁾, however, pulses of the exponential-recurrent type are for all practical

purposes of identical wave-form, and have advantages of their own from the theoretical point of view. Some of these advantages will be illustrated in the following sections by a more detailed consideration of the various possible types of modulation that may be applied to exponential-recurrent pulses.

The exponential-recurrent pulse as previously defined is

$$i = Ae^{BE \cos(\omega t + \phi)} \quad (17)$$

It appears, therefore, that intelligence might be transmitted on such a pulse by modulating either A , E , ω or ϕ . These four different types of modulation will now be examined.

2.1. Amplitude Modulation.

Let the signal it is desired to transmit be

$$e_s = E_s \sin pt. \quad (18)$$

If A in equation (17) is then made a linear function of e_s , the resultant pulse signal may be written

$$i = A(1 + q \sin pt)e^{BE \cos(\omega t + \phi)} \quad (19)$$

and this can be expanded into the following infinite series

$$i = A(1 + q \sin pt) \left\{ I_0(BE) + 2 \sum_{n=1}^{\infty} I_n(BE) \cos n(\omega t + \phi) \right\} \quad (20)$$

Further manipulation will clearly give terms in $\sin \{n(\omega t + \phi) + pt\}$ and $\sin \{n(\omega t + \phi) - pt\}$ amongst others, and it can be seen that this type of modulation is equivalent to amplitude modulation of a series of continuous carriers consisting of the fundamental and harmonics of the pulse repetition frequency ω . The resultant spectrum therefore contains the "carriers" $n\omega$ and the "sidebands" $(n\omega \pm p)$. It follows that the original intelligence contained in equation (18) is conveyed by all the pulse harmonics $n\omega$, and could in fact be obtained undistorted from any harmonic and its corresponding sidebands by simple rectification, provided that $2p \gg \omega$.

In some applications a number, say, m , of pulses of the type given by equation (17) are combined to give a composite wave-form in which adjacent pulses are displaced in time by $2\pi/m\omega$ seconds. In such a case all the carriers below the $(m+1)$ th disappear but the corresponding sidebands remain. This follows from the fact that

$$\sum_{r=1}^m \cos \left(\omega t + \frac{2\pi r}{m} + \phi \right) = 0, \quad (21)$$

when $n \leq m$.

Returning to the case of the single modulated recurrent pulse, the modulation signal of equation (18) may also be obtained from the modulated wave simply by eliminating all frequency components greater than p . A further method, at least of theoretical interest, is given by multiplying the modulated signal (19) by a recurrent pulse given by

$$i = Ce^{-BE \cos(\omega t + \phi)} \quad (22)$$

2.2. Combined Amplitude and Pulse Width Modulation.

Let the modulating signal be

$$e_1 = E_1 \cos pt,$$

and let the recurrent pulse be generated from

$$e_2 = E_2 \cos (\omega t + \phi).$$

Then if e_1 and e_2 are added together a form of modulation results as follows :

$$I = Ae^{B\{E_2 \cos(\omega t + \phi) + E_1 \cos pt\}}$$

$$\text{or } I = A \left\{ I_0(\text{BE}_2) + 2 \sum_{n=1}^{\infty} I_n(\text{BE}_2) \cos n(\omega t + \phi) \right\} \\ \times \left\{ I_0(\text{BE}_1) + 2 \sum_{n=1}^{\infty} I_n(\text{BE}_1) \cos n\pi t \right\},$$

or again,

$$I = A \left[I_0(\text{BE}_1) I_0(\text{BE}_2) + \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} I_m(\text{BE}_1) I_n(\text{BE}_2) \cos \{n(\omega t + \phi) + mpt\} \right], \quad (23)$$

where m and n are any integers or zero, with the restriction that both m and n may not be zero together. It is seen, therefore, that the resulting modulated signal contains a series of carriers, all harmonics of ω , each with sidebands corresponding to all harmonics of p .

The exact demodulation of such a signal may prove difficult, but by multiplying by an antiphased pulse similar to equation (22) a signal given by

$$I' = A' e^{BE_1 \cos pt} \quad (24)$$

can be obtained.

By making BE_1 small compared with unity the distortion present in the signal, equation (24), could probably be made negligible.

2.3. Phase and Frequency Modulation of the Pulse.

In equation (17) let

$$\phi = aE_2 \cos pt. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

Then

$$I = Ae^{BE_1 \cos(\omega t + aE_2 \cos pt)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

and this may be expanded as before

$$I=A\left\{I_0(BE_1)+2\sum_{n=1}^{\infty}I_n(BE_1)\cos n(\omega t+aE_2\cos pt)\right\}. \quad (27)$$

Regarding this as either phase or frequency modulation, it is seen that the spectrum consists of all harmonics of the pulse-repetition frequency ω modulated with increasing phase or frequency deviations as the order of the harmonic concerned increases: the number of sidebands involved at appreciable amplitude will thus tend to increase with the order of the harmonic of ω concerned, though their absolute values of course decrease with the factor $I_n(\text{BE}_a)$.

Demodulation of such a signal would thus have to rely on the lower-order harmonics of ω , as the sideband overlap of the higher orders must cause increasing distortion.

2.4. *Alternative Analysis of Crosstalk in a Multiplex System.*

Let $f(pt)$ be the signal it is desired to transmit, and let this be caused to amplitude modulate one of the recurrent pulses at the sending end. The output of the transmitter corresponding to this particular channel is then of the form

$$A.f(pt)e^{BE \cos \omega t} \dots \dots \dots (28)$$

As shown in the previous paper ⁽¹⁾, the exponential term may be expanded as an infinite series of terms in $\cos n\omega t$ whose coefficients involve the modified Bessel function $I_n(BE)$. Suppose, however, that the link between transmitter and receiver will pass only those harmonics of ω up to the m th without attenuation or phase distortion, and entirely suppresses all higher frequencies. The received signal corresponding to (28) is then given by

$$A.f(pt) \cdot \left\{ I_0(BE) + 2 \sum_{n=1}^m I_n(BE) \cos n\omega t \right\} \dots \dots \dots (29)$$

At the receiver this signal is demodulated by multiplying it by another pulse of the same phase and filtering out all high frequency components from the resultant. The unfiltered output is thus

$$A.f(pt) \cdot \left\{ I_0(BE) + 2 \sum_{n=1}^{\infty} I_n(BE) \cos n\omega t \right\} \left\{ I_0(BE) + 2 \sum_{n=1}^m I_n(BE) \cos n\omega t \right\} \dots \dots \dots (30)$$

The desired output after filtering must clearly consist of only the constant terms in the product of the two curly brackets in (30), and this gives

$$A.f(pt) \cdot \left\{ I_0^2(BE) + 2 \sum_{n=1}^m I_n^2(BE) \right\} \dots \dots \dots (31)$$

A similar argument to the above, applied to the case where the transmitted pulse has its phase shifted by an angle ϕ , but where the demodulated pulse is unchanged, shows that the filtered output in this case is given by

$$A.f(pt) \cdot \left\{ I_0^2(BE) + 2 \sum_{n=1}^m I_n^2(BE) \cos n\phi \right\} \dots \dots \dots (32)$$

Hence the cross-talk ratio between two channels separated by ϕ radians is

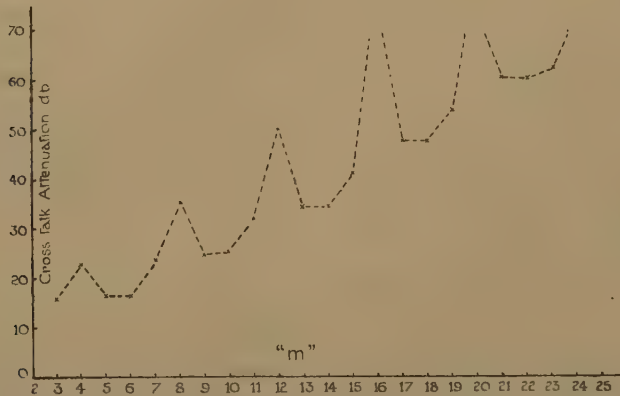
$$\text{C.T.R.} = \frac{I_0^2(BE) + 2 \sum_{n=1}^m I_n^2(BE)}{I_0^2(BE) + 2 \sum_{n=1}^m I_n^2(BE) \cos n\phi} \dots \dots \dots (33)$$

Making use of the approximation (8) of the previous paper⁽¹⁾ equation (32) above may be rewritten

$$\text{C.T.R.} = \frac{1 + 2 \sum_{n=1}^m e^{-n^2/BE}}{1 + 2 \sum_{n=1}^m (e^{-n^2/BE} \cdot \cos n\phi)} \quad (34)$$

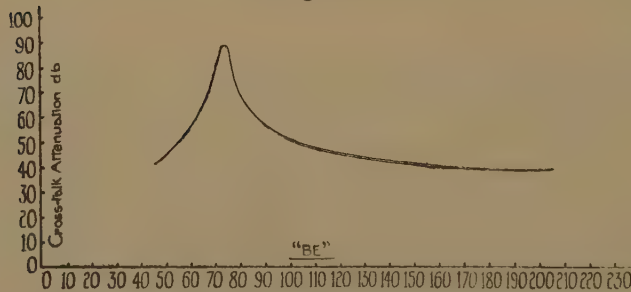
This expression has been used to compute cross-talk ratios for a series of values of m , that is, for a gradually increasing high frequency cut-off

Fig. 2.



The effect of H.F. cut-off on cross-talk, adjacent channels.
BE=100, $\phi=45^\circ$.

Fig. 3.



The effect of "B.E." upon adjacent channel cross-talk attenuation.
 $m=12$, $\phi=45^\circ$.

of the transmission link, and fig. 2 illustrates typical results. Fig. 3 illustrates the effect of BE upon the cross-talk attenuation for a transmission link having a perfect characteristic up to the m th harmonic and infinite attenuation for larger values. The information shown in the latter figure was computed before the close connection between the recurrent exponential pulse and the single probability pulse had been realized, consequently equation (33) was used in the computation.

The cross-talk produced by adjacent pulses being too close together, already referred to in section 1.4, may alternatively be derived from the equation of the exponential pulse.

As before, let the modulated output be

$$A.f(pt).e^{BE \cos \omega t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

then the output of the *n*th pulse from this is

$$A.f(pt).e^{BE \cos (\omega t + \overline{n-1}\phi)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

where ϕ is the angular distance between adjacent pulses. Demodulating as before, the signal contributed from the “ wanted ” pulse is

$$A.f(pt).e^{2BE \cos \omega t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

whilst the “ unwanted ” signal from the *n*th pulse is

$$A.f(pt)e^{BE\{\cos \omega t + \cos (\omega t + \overline{n-1}\phi)\}}$$

or

$$A.f(pt).e^{2BE \cos \{\omega t + (\frac{n-1}{2})\phi\} \cos (\frac{n-1}{2})\phi}. \quad . \quad . \quad . \quad . \quad . \quad (38)$$

When all frequencies not included in *f*(*pt*) are eliminated the ratio of the remaining signal, that is the cross-talk ratio, is

$$C.T.R.=\frac{I_0(2BE)}{I_0\left\{2BE \cos \left(\frac{n-1}{2}\right)\phi\right\}}. \quad . \quad . \quad . \quad . \quad . \quad (39)$$

Providing BE is greater than about 50 the Bessel function *I*₀(*x*) may be replaced by *e^x*/√2π*x* thus,

$$C.T.R.=\frac{e^{2BE\{1-\cos (\frac{n-1}{2})\phi\}}}{\sqrt{\cos \left(\frac{n-1}{2}\right)\phi}}. \quad . \quad . \quad . \quad . \quad . \quad (40)$$

For adjacent pulses *n*=2 and the cross-talk ratio is

$$\frac{e^{2BE\left(1-\cos \frac{\phi}{2}\right)}}{\sqrt{\cos \frac{\phi}{2}}}. \quad . \quad . \quad . \quad . \quad . \quad (41)$$

To correlate this expression with expression (3), use is made of equation (17) of the previous paper ⁽¹⁾, namely,

$$BE=\frac{1}{2\pi^2(FT)^2}, \quad . \quad . \quad . \quad . \quad . \quad (42)$$

and ϕ is assumed to be small. Then the cross-talk ratio becomes

$$e^{t_0^2/2T^2}, \quad . \quad . \quad . \quad . \quad . \quad (43)$$

or the cross-talk attenuation is

$$t_0^2/2T^2 \text{ nepers.} \quad . \quad . \quad . \quad . \quad . \quad (44)$$

Conclusions.

The addition and multiplication of the single probability function pulse has been demonstrated. As a further example of the theoretical use of this type of pulse the problem of cross-talk in multichannel time division communication systems has been considered. It has further been shown that the response of a network to the pulse can be determined by representing the frequency characteristic of the network by the sum of a number of exponential functions.

The recurrent exponential pulse and possible methods of modulating it have been dealt with, and the application of this type of pulse to cross-talk analysis of multichannel time division communication systems has been considered.

The writers wish to acknowledge the permission of the Engineer-in-Chief of the British Post Office to publish this paper.

Bibliography.

- (1) "Some Properties of a Special Type of Electrical Pulse," *Phil. Mag.* (Dec. 1943).
- (2) "Time-Division Multiplex System." Bennett, B. S. T. J. (April 1941).
- (3) "Untersuchungen an der wechselzeitigen Mehrfachübertragung (Multiplexübertragung)." Raabe, E. N. T. 1939. 19 pp. 213-228.

LIII. *Conditions of Failure and Factors of Safety in Members subject to Alternating Stresses.*

By S. V. SERENSEN, Member Acad. Sci. of the Ukraine *.

[Received March 28, 1944.]

EXPERIMENTAL data on the influence of static stresses are in agreement with Guest's⁽¹⁾ hypothesis, which gives the following conditions of plasticity expressed in terms of the principal stresses σ_1 and σ_3 :

$$\sigma_1 - \sigma_3 = 2\tau_s - \lambda(\sigma_1 + \sigma_3), \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where τ_s is the limit of flow in shear.

The conditions of plasticity are thus associated with volume deformations as well as with tangential stresses and shear deformations.

In the case of simultaneous application of bending stresses σ normal to the extension and torsional stresses τ tangential to it we have the relationships

$$\sigma = \sigma_1 + \sigma_3$$

and

$$\tau^2 = -\sigma_1 \sigma_3.$$

For these stresses expression (1) then becomes

$$(1 - \lambda)^2 \sigma^2 + 4\tau^2 + 4\lambda\tau_s \sigma = 4\tau_s^3. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

* Communicated by The British Council.

This is the equation of an ellipse with its centre on the σ -axis at a distance

$$d = \frac{2\lambda\tau_s}{1-\lambda^2} \text{ from the origin.}$$

The coefficient λ may be determined from the conditions of plasticity in simple tension and pure shear. Applying expression (1) for these cases we obtain

$$\lambda = 2 \frac{\tau_s}{\sigma_s} - 1,$$

where σ_s is the limit of flow in tension.

Using condition (1) for the case of fatigue failure with alternating stresses in a symmetrical cycle we obtain

$$(1-\lambda^2)\sigma^2 + 4\tau^2 + 4\lambda\tau_{-1}\sigma = 4\tau_{-1}^2, \quad \dots \quad (3)$$

where τ_{-1} is the fatigue limit in torsion with a symmetrical cycle, *i. e.* the coefficient of amplitude of the stress

$$\rho = \tau \frac{\tau_{\min}}{\tau_{\max}} = -1.$$

The coefficient λ should be

$$\lambda = 2 \frac{\tau_{-1}}{\sigma_{-1}} - 1,$$

where σ_{-1} is the fatigue limit in tension-compression (or bending) with a symmetrical cycle.

Equation (2) is then expressed thus

$$\left(\frac{\tau}{\tau_{-1}}\right)^2 + \left(\frac{\sigma}{\sigma_{-1}}\right)^2 \left(\frac{\sigma_{-1}}{\tau_{-1}} - 1\right)^2 + \frac{\sigma}{\sigma_{-1}} \left(2 - \frac{\sigma_{-1}}{\tau_{-1}}\right) = 1. \quad \dots \quad (4)$$

This relationship was confirmed experimentally by Gough and Pollard⁽³⁾ in fatigue tests on cast-iron with simultaneous bending and torsion. The hypothesis of Huber-Henki-Mises⁽⁴⁾ gives the conditions of plasticity only as a function of the tangential stresses on areas equally inclined to the directions of the principal stresses. For the state of plane stressing this condition is expressed by the formula

$$(\sigma_1 + \sigma_3)^2 - 3\sigma_1\sigma_3 = 3\tau_s^2. \quad \dots \quad (5)$$

If a relationship of this type is required to satisfy experimental conditions of plasticity in simple tension and pure shear it is expressed as⁽⁵⁾

$$(\sigma_1 + \sigma_3)^2 - \delta^2\sigma_1\sigma_3 = \sigma_s^2, \quad \dots \quad (6)$$

where $\delta = \frac{\sigma_s}{\tau_s}$.

Condition (6) may be written in the following form

$$\left(\frac{\sigma}{\sigma_s}\right)^2 + \left(\frac{\tau}{\tau_s}\right)^2 = 1. \quad \dots \quad (7)$$

Gough and Pollard⁽⁶⁾, working on fatigue test results for structural

In the following the amplitude coefficient for normal stresses will be denoted by ρ_σ and for tangential stresses by ρ_τ . In the expressions (3), (8) and (10) σ and τ will now be the maximum stresses in the cycle. In place of the fatigue limits in a symmetrical cycle there will figure the fatigue limits or static strength $\sigma_{\rho\sigma}$ and $\tau_{\rho\tau}$ for the corresponding asymmetric cycle. Thus from (3) we obtain

$$(1-\lambda^2)\sigma^2 + 4\tau^2 + 4\lambda\tau_{\rho\tau} = 4\tau_{\rho\tau}^2, \quad (11)$$

Fig. 1.

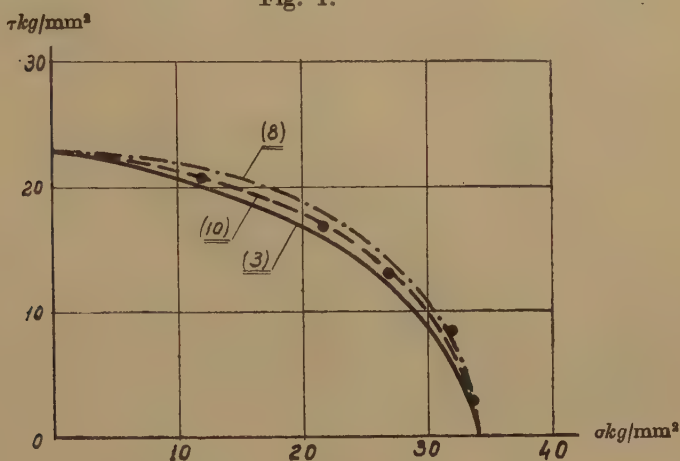
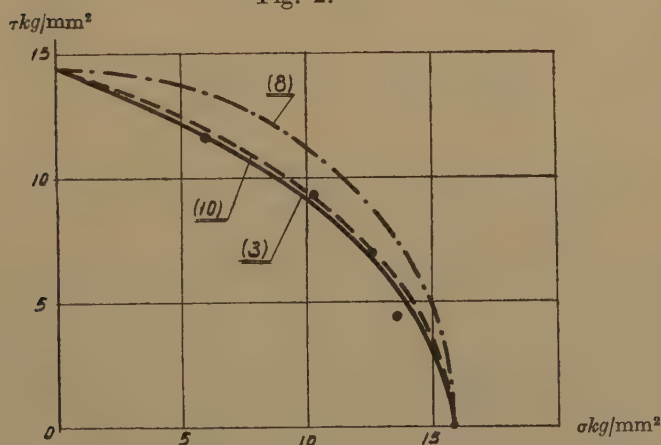


Fig. 2.



where

$$\lambda = 2 \frac{\tau_{\rho\tau}}{\sigma_{\rho\sigma}} - 1.$$

From expression (8)

$$\left(\frac{\sigma}{\sigma_{\rho\sigma}}\right)^2 + \left(\frac{\tau}{\tau_{\rho\tau}}\right)^2 = 1. \quad (12)$$

Using Goodman's linear relationships we arrive at the expression ⁽⁹⁾

$$\left(\frac{\sigma_m}{\sigma_s} + \frac{\sigma_v}{\sigma_{-1}}\right)^2 + \left(\frac{\tau_m}{\tau_s} + \frac{\tau_v}{\tau_{-1}}\right)^2 = 1, \quad . \quad . \quad . \quad (13)$$

where σ_m and τ_m are the static parts of the stress,
 σ_v and τ_v are their amplitudes.

In the principal stresses ⁽⁸⁾

$$\left(\frac{\sigma_{1m} + \sigma_{3m}}{\sigma_s} + \frac{\sigma_{1v} + \sigma_{3v}}{\sigma_{-1}}\right)^2 - \left(\frac{\sigma_s}{\tau_s}\right)^2 \left(\frac{\sigma_{1v}}{\sigma_{-1}} + \frac{\sigma_{1m}}{\sigma_s}\right) \left(\frac{\sigma_{2v}}{\sigma_{-1}} + \frac{\sigma_{2m}}{\sigma_s}\right) = 1. \quad (14)$$

For the general case of a plane stressed state characterized by stresses σ_x , σ_y and τ_{xy} we obtain ⁽⁸⁾

$$\left[\left(\frac{\sigma_{xv}}{\sigma_{-1}} + \frac{\sigma_{xm}}{\sigma_s}\right) + \left(\frac{\sigma_{yv}}{\sigma_{-1}} + \frac{\sigma_{ym}}{\sigma_s}\right)\right]^2 - \left(\frac{\sigma_s}{\tau_s}\right)^2 \left[\left(\frac{\sigma_{xv}}{\sigma_{-1}} + \frac{\sigma_{xm}}{\sigma_s}\right) \times \left(\frac{\sigma_{yv}}{\sigma_{-1}} + \frac{\sigma_{ym}}{\sigma_s}\right)\right] + \left(\frac{\tau_v}{\tau_{-1}} + \frac{\tau_m}{\tau_s}\right) = 1. \quad . \quad . \quad (15)$$

If we substitute $\frac{\sigma_s}{\tau_s} = \frac{\sigma_{-1}}{\tau_{-1}} = \sqrt{3}$, then expression (15) coincides with the conditions proposed by Marin ⁽¹⁰⁾ if the cycles of changes of normal and tangential stresses are similar (have the same coefficient of amplitude).

The following expression for the factor of safety with simultaneous bending and torsion follows from equation (13) ⁽¹¹⁾:

$$n = \frac{\sigma_s}{\sqrt{\left(1 + \frac{\sigma_v}{\sigma_m} \frac{\sigma_s}{\sigma_{-1}}\right)^2 \sigma_m^2 + \left(\frac{\sigma_s}{\tau_s}\right)^2 \left(1 + \frac{\tau_v}{\tau_m} \frac{\tau_s}{\tau_{-1}}\right)^2 \tau_m^2}},$$

$$= \frac{\sigma_{-1}}{\sqrt{\left(1 + \frac{\sigma_m}{\sigma_v} \frac{\sigma_{-1}}{\sigma_s}\right)^2 \sigma_v^2 + \left(\frac{\sigma_{-1}}{\tau_{-1}}\right)^2 \left(1 + \frac{\tau_m}{\tau_v} \frac{\tau_{-1}}{\tau_s}\right)^2 \tau_v^2}}. \quad . \quad . \quad (16)$$

If we substitute $\frac{\sigma_s}{\tau_s} = \frac{\sigma_{-1}}{\tau_{-1}} = \sqrt{3}$, then expression (16) is transformed into the formula proposed by Soderberg ⁽¹²⁾.

The actual strength for simple tension and simple torsion is best given experimentally by Smith's diagram rather than by Goodman's linear relationship. Therefore, in a strength equation of the type of (11) and (12) we may introduce the limits of strength σ_{ps} and τ_{ps} taken from Smith's diagram corresponding to the asymmetry of cycle (11).

Smith's diagram is conveniently constructed in schematic form from three experimental characteristics of σ_{-1} and σ_0 , where σ_0 is the fatigue limit with zero coefficient of amplitude. A diagram of this kind is constructed in fig. 3. It consists of two rectilinear parts. For the first of these the maximum stress in the cycle may be expressed as

$$\sigma_{\max} = \sigma_{md}(1 - \psi) + \sigma_{-1}, \quad . \quad . \quad . \quad (17)$$

where

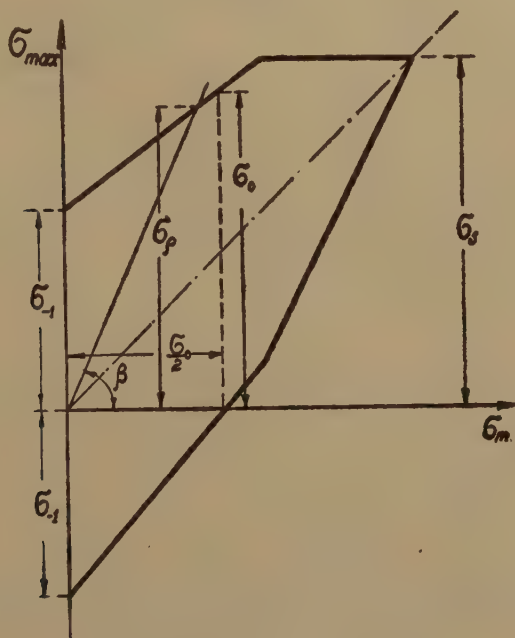
$$\psi = \frac{2\sigma_{-1} - \sigma_0}{\sigma_0}.$$

In the second one

$$\sigma_{\max} = \sigma_s.$$

Under the action of alternating stresses the strength is greatly influenced by stress concentrations, absolute dimensions, direction of grain and surface quality. With an asymmetric cycle these factors affect the amplitude of the failure stress. Reduction of this amplitude due to stress concentrations is characterized by an effective coefficient of concentration κ_σ (and correspondingly κ_τ), the change in strength due to the absolute dimensions by the coefficient ϵ and that connected with technical factors by the coefficient δ .

Fig. 3.



If normal stresses $\sigma_{\max} = \sigma_m + \sigma_v$ act in simple tension, then within the limits of the first portion of the Smith diagram (which is most commonly met with in calculations) the factor of safety is determined by the conditions of similarity of the cycles acting in the specimen and in the diagram. The influence of the concentrations and other factors is related to the stresses acting in the specimen, since the Smith diagram is usually constructed for smooth laboratory specimens. Taking into account expression (17) similarity of the cycles is expressed as follows :

$$\frac{\sigma_m}{\sigma_m + \frac{\kappa}{\epsilon \delta} \sigma_v} = \frac{\sigma_{m d}}{\sigma_{\max d}} = \frac{\sigma_m^n}{n \sigma_m (1 - \psi) + \sigma_1},$$

whence the factor of safety

$$n = \frac{\sigma_{-1}}{\phi\sigma_v + \psi\sigma_m}, \quad \dots \dots \dots (18)$$

where

$$\phi = \frac{\kappa}{\epsilon\delta}.$$

In expression (18) ϕ characterizes the effectiveness of the alternating part of the stress, ψ the effectiveness of the constant part. In machine specimens ϕ is usually considerably greater than ψ .

For the case of simultaneous application of bending (or tension) and torsion we obtain the following expression for the factor of safety from expression (12)

$$\frac{1}{n^2} = \left(\frac{\phi\sigma_v + \psi\sigma_m}{\sigma_{-1}} \right)^2 + \left(\frac{\phi\tau_v + \psi\tau_m}{\tau_{-1}} \right)^2. \quad \dots \dots \dots (19)$$

In the second (horizontal) part of the diagram the factor of safety is

$$n = \frac{\sigma_s}{\sigma_m + \phi\sigma_v}. \quad \dots \dots \dots (20)$$

If direct use is made of the Smith diagram without regard to its schematization the factor of safety is given by the following expression, which follows from expression (12):

$$\frac{1}{n^2} = \left(\frac{\sigma_m + \phi\sigma_v}{\sigma_{\rho\sigma}} \right)^2 + \left(\frac{\tau_m + \phi\tau_v}{\tau_{\rho\tau}} \right)^2. \quad \dots \dots \dots (21)$$

In this formula $\sigma_{\rho\sigma}$ is determined as the ordinate from the Smith diagram for a similar cycle. On the diagram this ordinate intersects a line drawn from the origin at an angle β and the σ_m axis. It is obvious from (18) that the angles β are

$$\tan \beta_\sigma = 1 + \frac{\kappa_\sigma}{\epsilon_\sigma \delta_\sigma} \frac{\sigma_v}{\sigma_m}, \quad \tan \beta_\tau = 1 + \frac{\kappa_\tau}{\epsilon_\tau \delta_\tau} \frac{\tau_v}{\tau_m}. \quad \dots \dots \dots (22)$$

According to Guest the following expression for the factor of safety is derived from expression (11):

$$\frac{1}{n} = \frac{1}{\tau_{\rho\tau}} \cdot \frac{\lambda}{2} \left(\sigma_{\max} + \frac{1}{\lambda} \sqrt{\sigma_{\max}^2 + 4\tau_{\max}^2} \right); \quad \dots \dots \dots (23)$$

in this case

$$\lambda = 2 \frac{\tau_{\rho\tau}}{\sigma_{\rho\tau}} - 1$$

$$\sigma_{\max} = \sigma_m + \phi\sigma_v, \quad \tau_{\max} = \tau_m + \phi\tau_v.$$

In expressions for the conditions of strength in complex stressed states it was assumed that the actions of the components of the stresses were of the same phase. Special experiments carried out by the author⁽¹³⁾ for carbon and alloy steels showed that the conditions of strength do not depend appreciably on the phase displacement.

Expressions for the factor of safety (19) and (23) reflect sufficiently completely the strength of the metal under asymmetric cycles of stress change. Condition (19) is in good agreement with experimental data

on steel. Condition (23) covers more satisfactorily a wider range of properties of metals, including cast-irons. These formulæ facilitate calculation of the strength of machine details under the influence of alternating stresses.

Literature.

- (1) Guest, J., *Philosophical Magazine* (July, 1900).
- (2) Guest, J., 'The Automobile Engineer,' vol. xxvii. p. 326 (1937).
- (3) Gough, J., and Pollard, H., *Proceedings of the Inst. Automobile Engineers*, vol. v. p. 96 (1937).
- (4) Mises, R., 'Nachrichten d. Gesellschaft d. Wissenschaft zu Göttingen,' 1913.
- (5) Lachmann, K., *Ingenieur-Archiv*, Heft 3, 1930.
- (6) Gough, J., and Pollard, H., *Proc. Inst. Mech. Eng.* p. 3 (Nov. 1935).
- (7) Serensen, S. V., "On the Criteria of Strength for a Volume Stressed State," *Comptes Rendus (Dokladi), Acad. Sci. of the Ukraine*, 1940, No. 2. (In Russian.)
- (8) Serensen, S. V., 'News (Izvestia) of the Technical Section of the Acad. Sci. U.S.S.R.' No. 7-8, 1938. (In Russian.)
- (9) Serensen, S. V., 'The Strength of Metals and Design of Machine Details,' State Scientific-Technical Publishers, 1937. (In Russian.)
- (10) Marin, J., 'Journal of Applied Mechanics,' No. 2 (1937).
- (11) Serensen, S. V., 'Bulletin (Vestnik) of the Metal Industry,' 1937, No. 7-8. (In Russian.)
- (12) Soderberg, R., *Transactions A.S.M.E.* 1930, APM. 52-2.
- (13) Serensen, S. V., 'Engineers' Symposium, Mechanics' Institute.' Acad. Sci. U.S.S.R. 1941, vol. i. Issue 1.

LIV. *Experiments on the Six Focal Lines due to Reflexions at the Surfaces of Two Plano-Cylindrical Lenses.*

By G. F. C. SEARLE, Sc.D., F.R.S., Cavendish Laboratory, Cambridge †.

[Received March 15, 1944.]

§ 1. *Introduction.*

THE six focal lines are beautiful to behold. The explanation of their positions and directions by general astigmatic theory is within the capacity of serious students of physics. If perfect images be sought, focal lines are unwelcome defects, but the truer view is that the convergence of rays to a point-image is a special case of an astigmatic pencil. Any rigorous treatment of geometrical optics must inevitably be based on astigmatic theory, and no apology is required for what is an essential element in the subject.

† Communicated by the Author.

§ 2. *Calculation of Principal Curvatures.*

In the present, or any other, astigmatic work, we find that, with the axis Ox normal to the final wave-front at the origin, the front is given, for points near O , by

$$x = \frac{1}{2}Sy^2 + Wyz + \frac{1}{2}Tz^2. \quad (1)$$

The quantities S , W , T , which are reciprocals of lengths, are known in terms of the optical constants of the system.

The two focal lines λ_1 , λ_2 , arising from the wave-front, are associated in direction and position with the planes and the magnitudes of the two principal curvatures of the front. We begin by finding the curvature R , the reciprocal of the radius of curvature, of the section of (1) made by the plane $z = y \tan \theta$, which contains the axis Ox . Using polar co-ordinates h , θ , in place of y , z , we have $y = h \cos \theta$, $z = h \sin \theta$, and then the curvature at O of the plane section is given by $R = 2x/h^2$. Thus

$$R = S \cos^2 \theta + 2W \sin \theta \cos \theta + T \sin^2 \theta \quad (2)$$

$$= \frac{1}{2}(S+T) + \frac{1}{2}(S-T) \cos 2\theta + W \sin 2\theta. \quad (3)$$

The curvature has a stationary value R_1 , at $\theta = \theta_1$, for variations of θ , when $(dR/d\theta)_1 = 0$, or when

$$(S-T) \sin 2\theta_1 = 2W \cos 2\theta_1, \quad (4)$$

or

$$\tan 2\theta_1 = 2W/(S-T). \quad (5)$$

Since $W \sin 2\theta_1 = \frac{1}{2}(S-T) \sin^2 2\theta_1 / \cos 2\theta_1$, we have, by (3),

$$R_1 = \frac{1}{2}(S+T) + \frac{1}{2}(S-T) \sec 2\theta_1. \quad (6)$$

In the range $-\frac{1}{2}\pi$ to $\frac{1}{2}\pi$ there is one value of the double angle which satisfies (5). This we take as $2\theta_1$. Then θ_1 is in the range $-\frac{1}{4}\pi$ to $\frac{1}{4}\pi$. Other double angles satisfying (5) are $2\theta_1 + \pi$ and $2\theta_1 - \pi$. One of these we select as $2\theta_2$. We may use either in (7), because

$$\sec(2\theta_1 + \pi) = \sec(2\theta_1 - \pi).$$

We have

$$R_2 = \frac{1}{2}(S+T) + \frac{1}{2}(S-T) \sec 2\theta_2, \quad (7)$$

where $\sec 2\theta_2 = -\sec 2\theta_1$ †.

For the planes of principal curvature, we may disregard repetitions of direction and need consider only θ_1 and $\theta_2 = \theta_1 \pm \frac{1}{2}\pi$. For $\theta_1 = -\frac{1}{4}\pi$, to $\theta_1 = 0$, we take $\theta_2 = \theta_1 + \frac{1}{2}\pi$. For $\theta_1 = 0$, to $\theta_1 = \frac{1}{4}\pi$, we take $\theta_2 = \theta_1 - \frac{1}{2}\pi$. To complete the specification, we take $\theta_2 = \frac{1}{2}\pi$ when $\theta_1 = 0$.

For any general value of θ , R is merely $1/\rho$, where ρ is the radius of curvature at O of the *plane curve* of the section of the surface by the plane $z = y \tan \theta$. The normals to the curve which lie in this plane are not, in general, normals to the surface, but they *are* normals to the surface if $\theta = \theta_1$ (or $\theta = \theta_2$). If, however, $\theta = \theta_1$, the section is a principal

† The method of finding R_1 , R_2 is a modification of that due to Dr. G. T. Bennett and used by me in Proc. Optical Convention, 1926, p. 815. He died 11 Oct. 1943. For many years he had been to me a faithful friend and mathematical helper.

section of the surface, and R_1 is one of the two principal curvatures R_1, R_2 of the surface at the origin.

The focal line λ_1 , corresponding to the principal curvature R_1 , is in the plane of R_2 , and λ_2 is in the plane of R_1 . Through any point P of an ordinary surface there pass two lines, s_1, s_2 , of curvature. In general, the curvature $1/\rho_1$ at a point in the principal plane containing s_1 depends on s_2 , the arc measured from P and $1/\rho_2$ depends on s_1 †. When, as in the experiments, $d\rho_1/ds_2=0$ and $d\rho_2/ds_1=0$ at O, the focal lines are perpendicular to the normal to the surface, which in the present case is Ox. The rays in the planes of R_1, R_2 pass through F_1, F_2 , where $OF_1=1/R_1, OF_2=1/R_2$. The algebraical sense of direction is, of course, implied, and R_1 is positive when F_1 is on the positive part of Ox.

We may say that the focal line λ_2 is conjugate to λ_1 , and that λ_1 is conjugate to λ_2 .

§ 3. Examples of Computation of Principal Curvatures.

The question as to which of two mutually perpendicular directions corresponds to a particular focal line selected from a pair can be perplexing. It may be of service to illustrate the application of (5), (6) and (7) by two examples. In the first, $S=0.04, W=-0.015, T=0.02 \text{ cm.}^{-1}$. By (5), $\tan 2\theta_1=-1.5$, and $2\theta_1=-56^\circ 19'$. We put $2\theta_2=2\theta_1+\pi=123^\circ 41'$. Then $\sec 2\theta_1=-\sec 2\theta_2=1.803$, and $\frac{1}{2}(S-T) \sec 2\theta_1=0.01803$. By (6), (7),

$$R_1=\frac{1}{2}(S+T)+\frac{1}{2}(S-T) \sec 2\theta_1=0.03+0.01803=0.04803 \text{ cm.}^{-1},$$

$$R_2=\frac{1}{2}(S+T)+\frac{1}{2}(S-T) \sec 2\theta_2=0.03-0.01803=0.01197 \text{ cm.}^{-1}.$$

Hence $OF_1=1/R_1=20.82$, $OF_2=83.54 \text{ cm.}$ Since $\theta_1=-28^\circ 10'$, and $\theta_2=61^\circ 50'$, the focal line λ_1 through F_1 makes angle $61^\circ 50'$, and the line λ_2 through F_2 makes $-28^\circ 10'$ with the plane Oxy. Rotation about O in the plane of yz is positive when counter-clockwise to an observer on the positive part of Ox.

In the second example, $S=0.02, W=0.011, T=-0.01 \text{ cm.}^{-1}$. By (5), $\tan 2\theta_1=0.73333$, and $2\theta_1=36^\circ 15'$. Then $\frac{1}{2}(S-T) \sec 2\theta_1=0.01860$. Hence by (6), (7),

$$R_1=0.005+0.01860=0.02360, \quad R_2=0.005-0.01860=-0.01360 \text{ cm.}^{-1}.$$

Hence $OF_1=1/R_1=42.37$, $OF_2=-73.53 \text{ cm.}$ The point F_2 is on the negative part of Ox.

Since $\theta_1=18^\circ 8'$, and $\theta_2=18^\circ 8'-\frac{1}{2}\pi=-71^\circ 52'$, the focal line λ_1 through F_1 makes $-71^\circ 52'$, and λ_2 through F_2 makes $18^\circ 8'$ with the plane Oxy.

§ 4. Geometry of System of Plano-Cylindrical Lenses.

Two plano-cylindrical lenses L, N (fig. 1) are placed with their plane faces D, J parallel and their convex surfaces E, G in contact at the point O, which is taken as origin. The lens L is the nearer to the reader.

† G. F. C. Searle, 'Experimental Optics,' § 270.

The generator OL of the convex surface of L is displaced through an acute angle 2β from the generator ON of N; this angle is positive in fig. 1. The radius of each cylinder is c . The axis Ox is normal to the plane faces D, J. The axis Oy bisects the acute angle and Oz the obtuse angle between the generators.

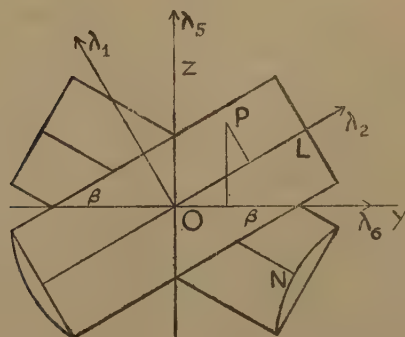
Let P_L be a point on the convex surface of L, with co-ordinates x_L, y, z . If p_L be the distance of P_L from the plane OxL, we have

$$p_L = -y \sin \beta + z \cos \beta. \quad (8)$$

For small values of p_L/c , the distance of P_L from the plane $x=0$ is $p_L^2/2c$. It is more convenient to use curvatures than to use radii, and we write $1/c=C$. Then, for the distance x_L of P_L from the plane Oyz, we have

$$x_L = \frac{1}{2} C p_L^2 = \frac{1}{2} C (-y \sin \beta + z \cos \beta)^2. \quad (9)$$

Fig. 1.



Let P_N be on the convex surface of N with co-ordinates x_N, y, z . If p_N be the distance of P_N from the plane OxN,

$$p_N = y \sin \beta + z \cos \beta. \quad (10)$$

The x -co-ordinate of P_N is negative, and

$$x_N = -p_N^2/2c = -\frac{1}{2} C p_N^2 = -\frac{1}{2} C (y \sin \beta + z \cos \beta)^2. \quad (11)$$

Let the thickness, measured parallel to Ox, of L at P_L be l_{yz} . Then $l_{yz} = l_0 - x_L$, where l_0 is the thickness along Ox, and thus

$$l_{yz} = l_0 - \frac{1}{2} C (y^2 \sin^2 \beta - yz \sin 2\beta + z^2 \cos^2 \beta). \quad (12)$$

Let the thickness of N at P_N be n_{yz} and along Ox be n_0 . Then $n_{yz} = n_0 + x_N$, and thus

$$n_{yz} = n_0 - \frac{1}{2} C (y^2 \sin^2 \beta + yz \sin 2\beta + z^2 \cos^2 \beta). \quad (13)$$

Let the thickness of the air-space between the curved surfaces be m_{yz} . Then

$$m_{yz} = x_L - x_N = C (y^2 \sin^2 \beta + z^2 \cos^2 \beta). \quad (14)$$

§ 5. *Optics of Cylindrical Lens System.*

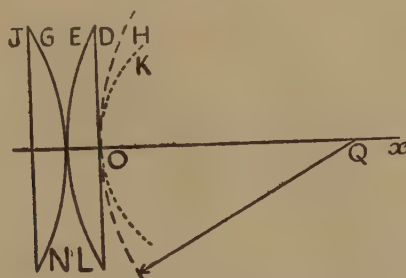
Light from a point Q (fig. 2) on Ox falls on the plane face D of lens L; Ox meets D in O and $QO = u$. The thickness of the system LN is

neglected in comparison with u . The spherical wave-front OH which meets D has radius u and curvature $U=1/u$. Some light is reflected at D. Other light is reflected at E, the curved surface of L, and at G and J, the curved and plane surfaces of N. We distinguish the four cases by the letters denoting the surfaces. To avoid confusion among cases (E), (G) and (J), we mark R, U, θ with subscripts 1, 2 for (E), with 3, 4 for (G), and with 5, 6 for (J). We consider the four cases in turn.

(D). The face acts as a plane mirror and gives a point-image of the point-source.

(E). The rays from Q, after refraction at D and reflexion at E, suffer a second refraction at D. The emergent rays have a wave-front OK, whose form we have to find.

Fig. 2.



With origin O in the plane face D, the incident front OH is, for small values of y, z ,

$$x_H = \frac{1}{2}U(y^2 + z^2). \quad (15)$$

Let the emergent front OK be

$$x_K = \frac{1}{2}Sy^2 + Wyz + \frac{1}{2}Tz^2. \quad (16)$$

The optical length [HK] of the path of a ray from H on OH to the corresponding point K on OK is independent of y and z and equals [OO], the optical length from O back to O. Hence [HK] - [OO] = 0. In this "thin" lens work, the effects on the optical distance [HK] of the very small angles made with Ox by the short parts of the ray are negligible.

With l_{yz} given by (12), and with [OO] = $2\mu l_0$, where μ is the refractive index, the optical equation is

$$[HK] - [OO] = x_H + x_K + 2\mu(l_{yz} - l_0) = 0.$$

Hence, by (15), (16),

$$\frac{1}{2}U(y^2 + z^2) + \frac{1}{2}Sy^2 + Wyz + \frac{1}{2}Tz^2 - \mu C(y^2 \sin^2 \beta - yz \sin 2\beta + z^2 \cos^2 \beta) = 0. \quad (17)$$

Since (17) holds for all (small) values of y, z ,

$$S = -U + 2\mu C \sin^2 \beta, \quad W = -\mu C \sin 2\beta, \quad T = -U + 2\mu C \cos^2 \beta. \quad (18)$$

By (5), $\tan 2\theta_1 = 2W/(S - T) = \tan 2\beta. \quad (19)$

We therefore put $\theta_1 = \beta$, $\theta_2 = \beta - \frac{1}{2}\pi$. Then, by (6), (7),

$$R_1 = -U + \mu C - \mu C = -U, \quad R_2 = -U + \mu C + \mu C = -U + 2\mu C. \quad (20), (21)$$

Since $\theta_1 = \beta$, the plane of the curvature R_1 is OxL (fig. 1). The focal line λ_1 , through the corresponding focal point F_1 , is normal to OxL . The line λ_2 , through F_2 , lies in OxL .

If Q be suitably placed, a focal line of the reflected beam will pass through Q . By analogy with the method of C. V. Boys for finding the radius of curvature of a face of a lens with spherical surfaces, we say that Q is then at a "Boys point."

The first Boys point B_1 is at $x = 1/U_1$, where $U_1 = R_1$. Since, by (20), $R_1 = -U_1$, we have

$$U_1 = 0, \quad \dots \dots \dots (22)$$

and thus B_1 is "at infinity." The point B_2 is at $x = 1/U_2$, where $U_2 = R_2$. Hence, by (21),

$$U_2 = \mu C, \quad \dots \dots \dots (23)$$

and thus $OB_2 = c/\mu$ and is positive.

(G). The rays from Q are reflected at G , the curved surface of N , and pass again through L . The optical equation is

$$[HK] - [OO] = x_H + x_K + 2\mu l_{yz} + 2m_{yz} - 2\mu l_0 = 0,$$

where m_{yz} is given by (14), and thus, by (12), (15), (16),

$$\begin{aligned} \frac{1}{2}U(y^2 + z^2) - \mu C(y^2 \sin^2 \beta - yz \sin 2\beta + z^2 \cos^2 \beta) \\ + 2C(y^2 \sin^2 \beta + z^2 \cos^2 \beta) = 0. \end{aligned} \quad (24)$$

Hence

$$\left. \begin{aligned} S &= -U - 2(2 - \mu)C \sin^2 \beta, & W &= -\mu C \sin 2\beta, \\ T &= -U - 2(2 - \mu)C \cos^2 \beta, & S - T &= 2(2 - \mu)C \cos 2\beta. \end{aligned} \right\} \quad (25)$$

$$\text{By (5),} \quad \tan 2\theta_3 = -\mu \tan 2\beta / (2 - \mu). \quad \dots \dots \dots (26)$$

Since $\tan 2\beta$ and $2 - \mu$ are positive, $\tan 2\theta_3$ is negative. We take $2\theta_3$ to be between $-\frac{1}{2}\pi$ and 0. Then $\sec 2\theta_3$ is positive. We have

$$\sec 2\theta_3 = \{1 + \mu^2 \tan^2 2\beta / (2 - \mu)^2\}^{\frac{1}{2}},$$

without ambiguity of sign. Hence, by (6), (7) and (25),

$$R_3 = -U - (2 - \mu)C + (2 - \mu)C \cos 2\beta \sec 2\theta_3 = -U - (2 - \mu - \omega)C, \quad (27)$$

$$R_4 = -U - (2 - \mu + \omega)C, \quad \dots \dots \dots (28)$$

where

$$\omega = \{(2 - \mu)^2 \cos^2 2\beta + \mu^2 \sin^2 2\beta\}^{\frac{1}{2}} = \{(\mu - 1)^2 + 1 - 2(\mu - 1) \cos 4\beta\}^{\frac{1}{2}}, \quad (29)$$

and ω is positive. For the lenses, $1 < \mu < 2$, and thus, as β goes from 0 to $\frac{1}{4}\pi$, ω goes from $2 - \mu$ to μ .

The angle θ_3 goes from 0 to $-\frac{1}{4}\pi$, and θ_4 from $\frac{1}{2}\pi$ to $\frac{3}{4}\pi$, as β goes from 0 to $\frac{1}{4}\pi$. The directions of λ_3 , λ_4 through F_3 , F_4 are given by θ_4 and θ_3 respectively.

The Boys point B_3 is at $x=1/U_3$, where $U_3=R_3$, and B_4 is at $x=1/U_4$. Hence, for B_3, B_4 ,

$$U_3=(\frac{1}{2}\mu-1+\frac{1}{2}\omega)C, \quad U_4=(\frac{1}{2}\mu-1-\frac{1}{2}\omega)C. \quad (30), (31)$$

Since $1<\mu<2$, U_3 goes from 0, when $\beta=0$, to $(\mu-1)C$, when $\beta=\frac{1}{4}\pi$, and is always positive. As β goes from 0 to $\frac{1}{4}\pi$, U_4 goes from $-(2-\mu)C$ to $-C$, and is always negative.

(J). The rays from Q are reflected at J, the plane face of N. The optical equation is

$$[HK]-[OO]=x_H+x_K+2\mu(l_{yz}+n_{yz})+2m_{yz}-2\mu(l_0+n_0)=0.$$

By (12), (13), no term in yz appears in $l_{yz}+n_{yz}$, and, by (14), (15), (16),

$$\frac{1}{2}U(y^2+z^2)+\frac{1}{2}Sy^2+Wyz+\frac{1}{2}Tz^2-2(\mu-1)C(y^2\sin^2\beta+z^2\cos^2\beta)=0. \quad (32)$$

Thus we have

$$\left. \begin{aligned} S &= -U + 4(\mu-1)C\sin^2\beta, & W &= 0, \\ T &= -U + 4(\mu-1)C\cos^2\beta, & S-T &= -4(\mu-1)C\cos^2\beta. \end{aligned} \right\} \quad (33)$$

By (5), $\tan 2\theta_5=2W/(S-T)=0$, and $\theta_5=0$, $\theta_6=\frac{1}{2}\pi$. By (6), (7),

$$R_5=-U+2(\mu-1)C-2(\mu-1)C\cos^2\beta=-U+4(\mu-1)C\sin^2\beta, \quad (34)$$

$$R_6=-U+2(\mu-1)C+2(\mu-1)C\cos^2\beta=-U+4(\mu-1)C\cos^2\beta. \quad (35)$$

Since $\theta_5=0$, the plane of the curvature R_5 is the plane Oxy , and λ_5 through F_5 is parallel to Oz . The line λ_6 is parallel to Oy , *i. e.* to the line bisecting the acute angle between the generators OL , ON (fig. 1).

The point B_5 is at $x=1/U_5$, where $U_5=R_5$; at B_6 , $U_6=R_6$. Hence

$$U_5=2(\mu-1)C\sin^2\beta, \quad U_6=2(\mu-1)C\cos^2\beta, \quad (36), (37)$$

and thus U_5 and U_6 are always positive. From (36), (37)

$$(U_5/U_6)^{\frac{1}{2}}=\tan\beta. \quad (38)$$

Since $0<\beta<\frac{1}{4}\pi$, U_6 exceeds U_5 and $OB_6<OB_5$.

If rays from Q, after passing through LN, meet a plane mirror P, which has Ox as a normal, any rays normal to J, the plane face of N, will, on reflexion at P, return along their own paths, which are those of rays reflected normally at J. Hence the Boys points for reflexion at P will agree with those for reflexion at J. If the P-normal be slightly inclined to the J-normal, the focal line due to P will be near and parallel to that due to J. This circumstance may be used to identify the focal lines due to J.

The intensity of a focal line due to J is diminished if a film of moisture, formed by breathing, be deposited on J.

The directions of $\lambda_1, \lambda_2, \lambda_5, \lambda_6$ are shown in fig. 1.

§ 6. Comparison of Results.

We may compare $U_3, U_4, \theta_3, \theta_4$, as derived from observations on λ_3, λ_4 , with the values computed from data obtained from $\lambda_1, \lambda_2, \lambda_5, \lambda_6$. By (23), $U_2=\mu C$, and, by (36), (37), $U_5+U_6=2(\mu-1)C$. Hence

$$C=\frac{1}{2}(2U_2-U_5-U_6), \quad \mu=U_2/C=2U_2/(2U_2-U_5-U_6). \quad (39), (40)$$

By (E), § 5, $\beta = \theta_1$, and $\beta = \frac{1}{2}\pi + \theta_2$. We use the mean

$$\beta = \frac{1}{2}(\theta_1 + \theta_2 + \frac{1}{2}\pi). \quad (41)$$

We use (40), (41) in (29) and compute ω . Taking C from (39) and using (30), (31), we compute U_3 , U_4 and compare them with the values derived from the observed reciprocals U_3^* , U_4^* , as obtained with the auxiliary lens A (§§ 9, 10).

The angle θ_3 , given, as in (26), by $\tan 2\theta_3 = -\mu \tan 2\beta/(2-\mu)$, is compared with that found from the directions of λ_3 , λ_4 .

Finally, the reciprocals U_5 , U_6 , as computed by (36), (37), are compared with those derived from the observed U_5^* , U_6^* .

§ 7. *Effects of Variations of β .*

The table shows how the positions and directions of the focal lines depend upon β ; the formulæ of § 5 were used. The lenses L, N are assumed to have $\mu = 1.52$ and radius $c = 1/C$. The table gives u_1/c , . . . , where u_1 , . . . are the distances of the B_1 , . . . -lines from LN. The

	$\beta = 0^\circ$	7.5°	15°	22.5°	30°	37.5°	45°
u_1/c	∞	∞	∞	∞	∞	∞	∞
u_2/c	0.66	0.66	0.66	0.66	0.66	0.66	0.66
θ_1	0°	7.5°	15°	22.5°	30°	37.5°	45°
u_3/c	∞	15.6	5.18	3.09	2.33	2.01	1.92
u_4/c	-2.08	-1.84	-1.49	-1.24	-1.10	-1.02	-1.00
θ_3	0°	-20.2°	-30.7°	-36.2°	-39.8°	-42.6°	-45°
u_5/c	∞	56.4	14.4	6.57	3.85	2.59	1.92
u_6/c	0.96	0.98	1.03	1.13	1.28	1.53	1.92
θ_5	0°	0°	0°	0°	0°	0°	0°

angles θ_1 , θ_3 , θ_5 give the directions of the planes of the principal curvatures corresponding to B_1 , B_3 , B_5 ; these are the directions of λ_2 , λ_4 , λ_6 . The angles θ_2 , θ_4 , θ_6 differ from θ_1 , θ_3 , θ_5 by $\frac{1}{2}\pi$.

§ 8. *The Gamma Points.*

When Q is at a Boys point B, a focal line of the reflected beam passes through B. The second focal line of the beam meets the axis in a point we call a Gamma Point and denote by Γ . Corresponding to a pair of Boys points, *e. g.* B_1 , B_2 , there are two Gamma points Γ_1 , Γ_2 , at distances $v_1 = 1/V_1$ and $v_2 = 1/V_2$ from the system. The positions of Γ_1 , Γ_2 are easily calculated when those of B_1 , B_2 are known.

Let $2t$ be the optical length of the path of a ray from the plane $x=0$ (as in fig. 2) back to that plane. Then, with a thin reflecting system, and with proper choice of axes, we have

$$2t = 2t_0 + My^2 + Nz^2. \quad (42)$$

The incident front from Q, at $x=u=1/U$ on Ox , is given by

$$x=\frac{1}{2}U(y^2+z^2).$$

The equation of the emergent front will have no term in yz ; it may be written

$$x=\frac{1}{2}Sy^2+\frac{1}{2}Tz^2.$$

The optical equation is

$$\frac{1}{2}U(y^2+z^2)+2t_0+My^2+Nz^2+\frac{1}{2}Sy^2+\frac{1}{2}Tz^2=2t_0.$$

Hence $S=-U-2M$, $T=-U-2N$ (43), (44)

The two principal curvatures of the emergent front are S , T . If Q be at the first of the two Boys points, $S_1=U_1$, and then, by (43), $S_1=U_1=-M$. We have $V_1=T_1$ and hence, by (44), $V_1=M-2N$. If Q be at B_2 , $T_2=U_2=-N$. Then $V_2=N-2M$. Hence

$$V_1=2U_2-U_1, \quad V_2=2U_1-U_2, \quad . \quad . \quad . \quad (45), (46)$$

and thus $3U_1=2V_2+V_1, \quad 3U_2=2V_1+V_2, \quad . \quad . \quad (47), (48)$

and $V_1+V_2=U_1+U_2, \quad V_1-V_2=3(U_2-U_1). \quad . \quad (49), (50)$

When lens A of fig. 3 is put into place, the constants t_0 , M , N in (42) are changed, and the distances of the Boys and Gamma points from the system become $u_1^* \dots, v_1^*, \dots$, with reciprocals $U_1^* \dots, V_1^* \dots$. The relations (45) to (50) hold for the latter reciprocals.

§ 9. Apparatus.

The plano-cylindrical lenses are spectacle lenses with a nominal focal length of 200 cm. They were "centred" and "edged" by an optician and are 4.4 cm. in diameter—as large as can be cut from the unedged lenses stocked by opticians. The distance between the plane faces D , J is 0.31 cm. For long focal lines, large lenses are needed. Three specks of gummed paper stuck, at angles of 120° , to one lens near its edge prevent contact between the lenses. Pressure produces a finite area of contact and distorts the lenses. The elliptical Newton's rings seen, by reflexion, with mono-chromatic light, have their centre on the shortest distance between the surfaces. If there be not contact, the centre of the pattern will alternate between bright and dark, as the angle of incidence is varied. With contact, the centre remains dark. The centre of the pattern can be brought near the centre of the lenses by suitable adjustment of the thicknesses of the specks of paper. The major axes of the ellipses lie along Oy (fig. 1).

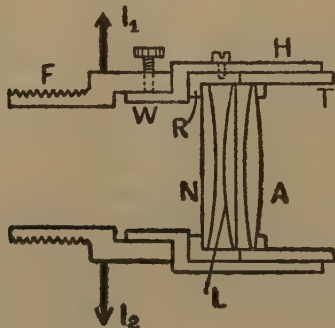
In assembling the system, the directions of the axes of the cylinders must be known. Each lens is laid with its cylindrical surface on a plane plate. The interference pattern seen with sodium light, ideally one of parallel straight lines, is generally one of very elongated ellipses, whose major axes are parallel to the axis of the cylinder. The direction of the axis may be marked on the cylindrical face by an ink line. Care should be taken that the axis of the lens L , which is to face the lamp, is displaced from that of N through a *positive* acute angle, as in fig. 1, so that, as

supposed in the theory, β may be positive. A little hard wax is run between the lenses near the specks of paper by aid of a hot wire.

The lenses L, N are mounted with soft wax in a ring-seating R (fig. 3), secured by screws in the holder H. This has a shank W, which is held in a fitting F screwed into the threaded opening in the ball bearing clutch of a free-wheel cycle. The internal pawls are removed. The toothed wheel of the clutch is fixed to a block carrying a circular scale; the block is borne on a carriage on an optical bench. The fitting F carries pointers I_1, I_2 moving over the circular scale. The large opening through the clutch allows the passage of a wide beam of light.

The Boys point B_4 is "virtual" for all values of β ; although the other five are "real," some may be beyond the range of the bench for some values of β . The difficulty is met by an auxiliary thin spherical double convex lens A of focal length $f=50$ cm.. It is fitted into a ring-seating at the end of the tube T, which slides into H. The lens A is

Fig. 3.



0.2 cm. thick and is nearly in contact with the face D of lens L †. To a first approximation, the lenses A, L, N may be treated as coincident.

The lens A changes the positions of the Boys' points from B_1, \dots to B_1^*, \dots . The directions of the focal lines are not changed, and we still denote the lines by λ_1, \dots . The points B_1 and B_1^* are conjugate with respect to A, and similarly for B_2, B_2^*, \dots .

The luminous point Q is the short filament of a 2-volt pea lamp (figs. 4, 5) supplied with current by a bell transformer taking 220 volts from the A.C. mains and yielding about 3 volts. A suitable resistance is in series with the lamp. The lamp and a frame F for a screen S of ground glass are fixed to the upper side of a horizontal plate P. The lower side carries a block bored to suit the bench fittings and provided with a clamping screw. A compact unit is obtained if the transformer be attached to the lower side of P. The ground glass slides in grooves in

† In fig. 3, for clearness, the thicknesses of L, N, A and the spaces between them are exaggerated.

F and is removable. The ground face is towards LN and its plane passes through Q.

The device is convenient for optical experiments, and particularly for the location of Boys points. It is true that the image on the screen does not actually coincide with Q, but it need not be more than two cm. from Q. If the distance of Q from the reflecting surface be more than 30 cm., the slight obliquity causes little error. If the ground glass be removed, the reflected rays can be received on a screen, borne on a separate carriage, at any desired distance from Q.

A square M resting on P gives a vertical direction to which the focal lines can be adjusted.

§ 10. Experimental Details.

The system LN is set so that a line through the centre of L and parallel to the bench meets the screen about one cm. from Q. The axis of LN is nearly parallel to the bench. The lens A is put in position. If the

Fig. 4.

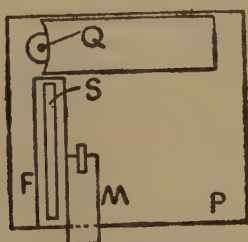
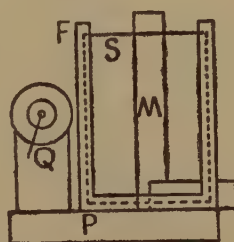


Fig. 5.



glowing lamp Q be now moved along the bench, the six focal lines will come, one after another, into view on the screen. Although lens A may not be needed for B_2 , it is convenient to use it throughout the observations on the focal lines.

The holder H (fig. 3) is adjusted in the fitting F so that the pointers I_1, I_2 are, as nearly as may be, vertical when λ_5 is vertical. This line, as also λ_6 , is due to reflexion at J. The two lines are identified by aid of a plane mirror P, as described in (J), § 5. The distances of λ_5, λ_6 from A are u_5^*, u_6^* , and $u_5^* > u_6^*$; they are found from the bench readings of the lamp carriage by use of a distance rod.

A rough preliminary survey is profitable. The line λ_5 is set and is kept vertical, and u_5^* is found approximately. The lamp is then moved to focus the other lines on the screen. Their distances u_1^*, \dots from A and rough values, with their signs, of their position angles θ_1, \dots relative to λ_5 are noted, as in the table. The angles may be estimated by aid of a transparent protractor resting on the plate P (fig. 5).

The observations proper can now be made. Line λ_5 is focused and is set vertical by aid of the square; the bench reading of the lamp and the readings of I_1, I_2 are taken. The settings are repeated two or three times and the means for I_1, I_2 are found; the mean bench reading

If Γ , Γ^* , at distances $1/V$, $1/V^*$ from L , be the corresponding Gamma points, they will be at S , S^* . Hence

$$V=V^*-1/f. \quad . \quad . \quad . \quad . \quad . \quad (52)$$

§ 11. Application of Newton's Rings.

The radius $c=C^{-1}$ cm. of the cylinders and the angle β can be found from the elliptical rings formed between the cylindrical faces. Monochromatic light of wave-length λ is used at normal incidence. By (14), the thickness, $h=m_{yz}$, of the air space at y , z is given by

$$ch=y^2/\operatorname{cosec}^2\beta+z^2/\sec^2\beta,$$

and hence the line of constant distance h is an ellipse of semi-axes $(ch)^{\frac{1}{2}}|\operatorname{cosec}\beta|$ and $(ch)^{\frac{1}{2}}|\sec\beta|$. If, as in the experiment, $0<\beta<\frac{1}{4}\pi$, $\operatorname{cosec}^2\beta>\sec^2\beta$, and the major axis lies along Oy in fig. 1. Let $2a_n$, $2b_n$ be the major and minor diameters of the n th dark ellipse, counting outwards from the centre, and let h_n be the corresponding thickness. Then $a_n=(ch_n)^{\frac{1}{2}}|\operatorname{cosec}\beta|$, and

$$4c(h_m-h_n)\operatorname{cosec}^2\beta=4(a_m^2-a_n^2).$$

But $h_m-h_n=\frac{1}{2}(m-n)\lambda$, and thus

$$2c\lambda=4\left(\frac{a_m^2-a_n^2}{m-n}\right)\sin^2\beta.$$

Observations are made with a fixed value—say 6—for $m-n$ and with $n=4, 5, \dots$. The rings for which $n<4$ are diffuse. The mean of $4(a_m^2-a_n^2)/(m-n)$ for the series of values of n is denoted by D^2 , and the similar mean of $4(b_m^2-b_n^2)/(m-n)$ is E^2 . Then, taking averages, we have

$$\sin^2\beta=2c\lambda/D^2, \quad \cos^2\beta=2c\lambda/E^2.$$

Hence $\tan\beta=E/D$, (53)

and $c=\frac{D^2E^2}{2(D^2+E^2)\lambda}$ (54)

§ 12. Practical Example.

The observations on focal lines were made by Mr. C. G. B. Garrett. Auxiliary lens A was used. The table gives distances $u_1^*, \dots v_1^*, \dots$ of the Boys and Gamma points from A and their reciprocals $U_1^*, \dots V_1^*, \dots$, and also the angles between the focal lines $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_6$ and the line λ_5 . The values of U_1, \dots are calculated by (51). The focal length of A equals u_1^* , viz. 49.62 cm.

The great distance v_3^* was found by aid of a subsidiary lens S of focal length 101.5 cm. placed 100 cm. from lens A . The line was then observed 98.1 cm. from S . Hence $v_3^*=3028$ cm.

By (39), $C=0.00953$ cm.⁻¹; the radius, $c=C^{-1}$, of the cylinders is 104.9 cm. By (40), $\mu=1.5152$. By (41), $\beta=26.43^\circ$. By (29),

$$\omega=\{(\mu-1)^2+1-2(\mu-1)\cos 4\beta\}^{\frac{1}{2}}=1.2428,$$

By (30), (31),

$$U_3 = (\tfrac{1}{2}\mu - 1 + \tfrac{1}{2}\omega)C = 0.003612, \quad U_4 = (\tfrac{1}{2}\mu - 1 - \tfrac{1}{2}\omega)C = -0.008232 \text{ cm.}^{-1}$$

By (26), $\tan 2\theta_3 = -\mu \tan 2\beta/(2-\mu) = -4.1265$, and $\theta_3 = -38.18^\circ$.

The mean derived from readings for λ_3, λ_4 gives

$$\theta_3 = \tfrac{1}{2}(-37.86^\circ + 52.26^\circ - 90^\circ) = -37.80^\circ.$$

Line.	θ deg.	u^* cm.	100 U^* cm. ⁻¹	100 U cm. ⁻¹	v^* cm.	100 V^* cm. ⁻¹
λ_1	26.93	49.62	2.015	0.000	20.34	4.916
λ_2	-64.07	28.91	3.459	1.444	169.8	0.589
λ_3	-37.86	42.00	2.381	0.366	3028	0.033
λ_4	52.26	83.41	1.199	-0.816	27.99	3.573
λ_5	- 0.08	45.41	2.202	0.187	29.20	3.425
λ_6	89.91	35.59	2.810	0.795	61.77	1.619

We can compute U_5, U_6 by (36), (37) and find

$$U_5 = 0.001945 \text{ cm.}^{-1}, \quad U_6 = 0.007874 \text{ cm.}^{-1}.$$

By § 8, and by (51), (52), $100V_1^* = 100(2U_2^* - U_1^*)$. This value of $100V_1^*$ and similar results computed for the other Gamma points are given below :—

Gamma point.	1	2	3	4	5	6
100 V^* cm. ⁻¹ .	4.903	0.571	0.017	3.563	3.418	1.594
v^* cm.	20.40	175.1	5882	28.07	29.26	62.74

Each of the computed values $100V_1^*, \dots$ is slightly less than the observed value in the previous table. The average difference is 0.015 dioptre, the power of a lens of focal length 67 metres. The small differences may be due to the neglected thickness of the system or to slight imperfections in the plane faces of L and N.

An image of the filament was found at the same distance, 169.8 cm., from A as F_2^* ; it is due to reflexion at D (fig. 2). It and the filament at 28.91 cm. from A are conjugate with respect to the "thick mirror" † formed by A and the face D. We have $\tfrac{1}{2}(169.8^{-1} + 28.91^{-1}) = 0.02024 \text{ cm.}^{-1}$ against 0.02015 ($=U_1^*$), which corresponds to the "centre" of the "thick mirror."

† G. F. C. Searle, 'Experimental Optics,' § 117.

Sodium light of mean $\lambda=5.893 \times 10^{-5}$ cm. was used by G. F. C. Searle in finding c and β by Newton's rings (§ 11). With values 4, 5, . . . 9 for n and $m-n=6$, it was found that $D^2=0.066849$, $E^2=0.015507$ cm.². Then, by (53),

$$\tan \beta = E/D = 0.48163, \quad \beta = 25.71^\circ.$$

By (5),
$$c = \frac{1}{2} D^2 E^2 / (D^2 + E^2) \lambda = 106.8 \text{ cm.}$$

LV. On the Measurement of the Residual Parameters of a "Q" Meter.

By W. F. LOVERING *.

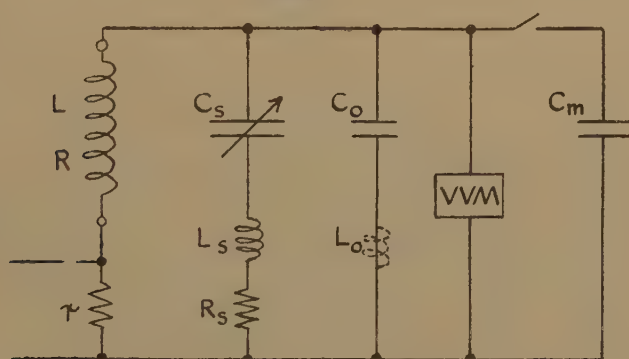
[Received February 18, 1944.]

IN a recent article J. C. Simmonds described a method of measuring the self-inductance and resistance of the tuning condenser of a "Q" meter by measurements made on the "Q" meter itself. The measurement can be carried out only if the calibration of the tuning condenser alone is known, but in general the low frequency calibration includes all stray capacities such as that of the valve voltmeter, and the formulæ given are invalid. The unknown stray capacity and the other parameters may be measured by extensions of the original method.

Measurement of Residual Inductance and Stray Capacity.

For this purpose the circuit of the "Q" meter may be represented by fig. 1, where C_s is the capacity of the variable tuning condenser (including

Fig. 1.



any stray capacity directly associated with it), L_s and R_s are the self-inductance and resistance of C_s , and C_0 is the remaining stray capacity. L is a coil of resistance R , r the injector resistance and C_m the unknown capacity. As in the original paper, it will be assumed that all measurements are made at frequencies much lower than the natural resonance

* Communicated by Professor L. F. Bates.

frequency of the tuning condenser, so that $\omega^2 L_s C_s$ is always very small compared with unity.

The procedure is as follows:—the circuit is tuned to resonance as indicated by maximum deflection of the valve voltmeter and the capacity reading C_1 is noted. C_1 is given from the low frequency calibration and is equal to $C_{s1} + C_0$. In this condition:—

$$\frac{1}{\omega^2 L} = C_{s1}(1 + \omega^2 L_s C_{s1}) + C_0 = C_1 + \omega^2 L_s C_{s1}^2, \quad \dots \quad (i)$$

where C_{s1} is the capacity of the tuning condenser at this setting.

The unknown capacity C_m is now connected across the circuit and C_s readjusted to a new value C_{s2} for resonance, and the capacity reading $C_2 = C_{s2} + C_0$ is noted. Then:—

$$\frac{1}{\omega^2 L} = C_m + C_{s2}(1 + \omega^2 L_s C_{s2}) + C_0 = C_m + C_2 + \omega^2 L_s C_{s2}^2, \quad \dots \quad (ii)$$

from (i) and (ii)

$$\begin{aligned} C_1 - C_2 &= C_m - \omega^2 L_s (C_{s1}^2 - C_{s2}^2) \quad \dots \quad (iii) \\ &= C_m - \omega^2 L_s (C_{s1} - C_{s2})(C_{s1} + C_{s2}) \\ &= C_m - \omega^2 L_s (C_1 - C_2)(C_1 + C_2 - 2C_0), \end{aligned}$$

$$\text{whence} \quad (C_1 - C_2)(1 - 2\omega^2 L_s C_0) = C_m - \omega^2 L_s (C_1^2 - C_2^2). \quad \dots \quad (iv)$$

The measurement is now repeated at the same frequency with other values of C_1 by using different coils, and a series of values for $C_1 - C_2$ and $C_1^2 - C_2^2$ obtained. A graph showing $C_1^2 - C_2^2$ plotted against $C_1 - C_2$ should be a straight line having a slope S , where

$$S = \frac{-(1 - 2\omega^2 L_s C_0)}{\omega^2 L_s},$$

$$\text{or} \quad S = 2C_0 - \frac{1}{\omega^2 L_s}. \quad \dots \quad (v)$$

The whole series of measurements is now repeated at other frequencies. S being found as above at each frequency and a graph plotted of S against $\frac{1}{\omega^2}$. This graph should also be a straight line, the intercept on the S -axis being $2C_0$ and the slope of the line $\frac{-1}{L_s}$, whence both C_0 and L_s may be found.

This measurement is unaffected by any inductance L_0 associated with C_0 (shown dotted in fig. 1), for in this case equation (i) becomes

$$\frac{1}{\omega^2 L} = C_{s1}(1 + \omega^2 L_s C_{s1}) + C_0(1 + \omega^2 L_0 C_0), \quad \dots \quad (vi)$$

and equation (ii) becomes

$$\frac{1}{\omega^2 L} = C_m + C_{s2}(1 + \omega^2 L_s C_{s2}) + C_0(1 + \omega^2 L_0 C_0), \quad \dots \quad (vii)$$

and from (vi) and (viii)

$$C_1 + \omega^2 L_s C_{s1}^2 + \omega^2 L_0 C_0 = C_2 + \omega^2 L_s C_{s2}^2 + \omega^2 L_0 C_0 + C_m,$$

or

$$C_1 - C_2 = C_m - \omega^2 L_s (C_{s1}^2 - C_{s2}^2),$$

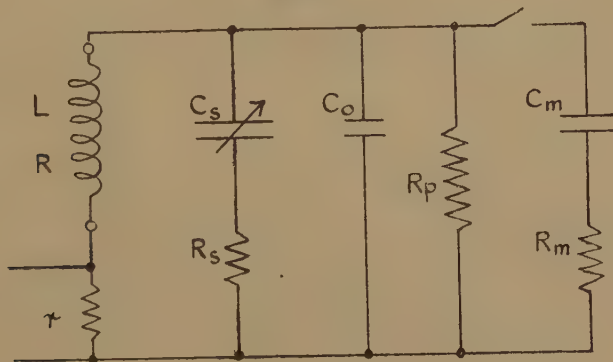
which is identical with equation (iii).

The effect of L_0 is to cause the total tuning capacity to be increased by a small quantity $\omega^2 L_0 C_0^2$. At frequencies below 15 Mc/sec. this term is usually negligible, since both L_0 and C_0 are small. At higher frequencies the representation of circuit inductances and capacities by "lumped" quantities is not justified and any attempt at a solution must use methods appropriate to transmission lines. In general, therefore, it is permissible to ignore L_0 within the range of frequencies to which the above analysis is applicable.

Determination of Residual Resistances.

For this purpose the circuit may be redrawn as in fig. 2, where C_s is the effective capacity of the tuning condenser corrected for the presence of L_s , R_p is any damping resistance in parallel with C_0 and R_m is the resistance of C_m .

Fig. 2.



Determination of R_s .

The circuit is tuned to resonance and the circuit step-up Q_1 noted, then :—

$$\frac{\omega C_1}{Q_1} = \frac{R+r}{\omega^2 L^2} + R_s \omega^2 C_{s1}^2 + \frac{1}{R_p}, \quad \dots \dots \dots \text{(viii)}$$

where

$$C_1 = C_0 + C_{s1},$$

C_{s1} being the corrected value of C_s

$$= C_1^1 + \omega^2 L_s (C_1^1 - C_0)^2,$$

where C_1^1 is the actual capacity reading as given by the low frequency calibration.

C_m is now connected, the circuit retuned to resonance, and the new Q value Q_2 and the new capacity reading C_2^1 noted. In this case :—

$$C_2 + C_m = C_1,$$

where C_2 is the corrected value of C_2^1 and

$$\frac{\omega(C_2 + C_m)}{Q_2} = \frac{R + r}{\omega^2 L^2} + R_s \omega^2 C_{s2}^2 + \frac{1}{R_p} + \omega^2 C_m^2 R_m, \quad \dots \quad (\text{ix})$$

subtracting (viii) from (ix) and substituting $C_1 = C_2 + C_m$ and $C_m = C_{s1} - C_{s2}$,

$$\omega C_1 \left(\frac{1}{Q_2} - \frac{1}{Q_1} \right) = R_m \omega^2 C_m^2 - \omega^2 R_s (C_{s2} + C_{s1}) \cdot C_m,$$

whence
$$\frac{C_1}{C_m} \left(\frac{Q_1 - Q_2}{Q_1 Q_2} \right) = R_m \omega C_m - \omega R_s (C_{s2} + C_{s1}). \quad \dots \quad (\text{x})$$

The measurement is repeated with different values of C_1 and a graph plotted showing $\frac{C_1}{C_m} \left(\frac{Q_1 - Q_2}{Q_1 Q_2} \right)$ plotted against $C_{s2} + C_{s1}$. This graph should be a straight line, the slope being $-\omega R_s$ and the intercept $R_m \omega C_m$.

It will be noted that this measurement takes no account of the resistance R_p , the possible presence of which was ignored in the original paper.

Determination of R_p .

For this measurement two coils L_A and L_B and an auxiliary variable condenser C_x are required. The coils should preferably be enclosed in high conductivity cans so that they have no external field. The coil L_B is tuned to resonance in the circuit and the Q value Q_1 and the corrected capacity C_1 noted; then:—

$$\frac{\omega C_1}{Q_1} = \frac{1}{Q_B \omega L_B} + R_s \omega^2 C_{s1}^2 + \frac{1}{R_p} + r \omega^2 C_1^2, \quad \dots \quad (\text{xi})$$

where Q_B is the Q value of L_B .

Coil L_A is now tuned to resonance by the condenser C_x and connected in parallel with the tuning condenser. If L_A and C_x are correctly tuned the setting of C_1 for resonance will be unaltered, but a new Q value Q_2 will be noted. In this condition:—

$$\frac{\omega C_1}{Q_2} = \frac{1}{Q_B \omega L_B} + R_s \omega^2 C_{s1}^2 + \frac{1}{R_p} + r \omega^2 C_1^2 + \frac{Q_A + Q_x}{Q_A Q_x \omega L_A}, \quad \dots \quad (\text{xii})$$

where Q_x is the Q value of C_x , which must be determined by a separate measurement, and Q_A is the Q value of the coil L_A .

From (xi) and (xii)
$$\omega C_1 \left(\frac{1}{Q_2} - \frac{1}{Q_1} \right) = \frac{Q_x + Q_A}{Q_A Q_x \omega L_A}. \quad \dots \quad (\text{xiii})$$

The coil L_B is now removed from the "Q" meter and L_A used in its place and tuned to resonance, the values C_3 of the tuning condenser and the Q value Q_3 being noted. In this case:—

$$\frac{\omega C_3}{Q_3} = \frac{1}{Q_A \omega L_A} + R_s \omega^2 C_{s3}^2 + \frac{1}{R_p} + r \omega^2 C_3^2. \quad \dots \quad (\text{xiv})$$

From (xiii)
$$\frac{1}{Q} = \frac{\omega L_A \omega C_1 (Q_1 - Q_2) Q_x - Q_1 Q_2}{Q_1 Q_2 Q_x},$$

whence

$$\frac{\omega C_3}{Q_3} = \frac{\omega L_A \omega C_1 (Q_1 - Q_2) Q_x - Q_1 Q_2}{Q_1 Q_2 Q_x \omega L_A} + R_s \omega^2 C_{s3}^2 + \frac{1}{R_p} + r \omega^2 C_3^2 \quad (\text{xv})$$

$$= \frac{\omega C_1 (Q_1 - Q_2) Q_x - Q_1 Q_2 \omega C_3}{Q_1 Q_2 Q_x} + R_s \omega^2 C_{s3}^2 + \frac{1}{R_p} + r \omega^2 C_3^2, \quad (\text{xvi})$$

since $\omega L_A = \frac{1}{\omega C_3}$,

whence R_p can be found since all the other quantities in the equation are known.

If L_A has an appreciable self-capacity this must be determined from measurements of the apparent inductance at two frequencies. If the self-capacity is C_A the apparent inductance is $\frac{L_A}{1 - \omega^2 L_A C_A}$, and this value must be substituted for L_A in equation (xv). Since

$$\frac{\omega L_A}{1 - \omega^2 L_A C_A} = \frac{1}{\omega C_3}, \quad \omega L_A = \frac{1}{\omega (C_A + C_3)}$$

and equation (xvi) becomes

$$\frac{\omega C_3}{Q_3} = \frac{\omega C_1 (Q_1 - Q_2) Q_x - Q_1 Q_2 \omega (C_A + C_3)}{Q_1 Q_2 Q_x} + R_s \omega^2 C_{s3}^2 + \frac{1}{R_p} + r^2 \omega^2 C_3^2, \quad (\text{xvii})$$

from which R_p can be found by substitution.

Summary.

Measurements are described by which J. C. Simmonds' method for determination of the residual parameters of a "Q" meter may be extended to take account of stray capacity and resistance not associated directly with the variable condenser, a knowledge of the low frequency calibration of the tuning condenser alone being unnecessary.

Reference.

Simmonds, J. C., *Phil. Mag.* vol. xxxiv. (1943).

LVI. Note on Green's Functions in the Theory of Heat Conduction.

By ARNOLD N. LOWAN*.

[Received October 14, 1943.]

It is a well-known fact that the solution of the most general problem in heat conduction reduces to the construction of the "Green's Function" appropriate to the given domain and to the nature of the boundary conditions†.

* Communicated by the Author.

† See article 80, "Mathematical Theory of the Conduction of Heat in Solids," H. S. Carslaw (1921).

By suitable superposition of the solutions corresponding to sources and sinks, it is possible to obtain the Green's Functions for domains bounded by planes which are either kept at 0° or are impervious to heat.

It is the object of this paper to derive the expressions of the Green's Functions for certain domains for which, to the author's knowledge, these expressions are not as yet given in the literature.

$$\text{Domain A} \begin{cases} 0 < x < \infty, \\ 0 < y < \infty. \end{cases}$$

Case 1. $G=0$ for $x=0$; $G=0$ for $y=0$.

We start with the solution

$$u_1 = u_1(x, y, \xi, \eta; t) = \frac{1}{4kt} \exp \left[-\frac{(x-\xi)^2 + (y-\eta)^2}{4kt} \right],$$

satisfying the condition of a source at the point (ξ, η) . With this source we associate the two sinks u_2 and u_3 at the points $(-\xi, \eta)$ and $(\xi, -\eta)$ and the source u_4 at $(-\xi, -\eta)$.

It is readily seen that the expression $u_1 + u_2 + u_3 + u_4$ satisfies the given boundary conditions. With the aid of the identity

$$\int_0^\infty e^{-a^2 x^2} \cos 2bx \, dx = \frac{\sqrt{\pi}}{2a} e^{-b^2/a^2},$$

the solution (1) may be written in the form

$$u_1 = \frac{1}{\pi^2} \int_0^\infty \int_0^\infty \exp[-k(\alpha^2 + \beta^2)t] \cos \alpha(x-\xi) \cdot \cos \beta(y-\eta) \, d\alpha \, d\beta.$$

Since

$$u_2(x, y, \xi, \eta; t) = -u_1(x, y, -\xi, \eta; t),$$

$$u_3(x, y, \xi, \eta; t) = -u_1(x, y, \xi, -\eta; t),$$

$$u_4(x, y, \xi, \eta; t) = u_1(x, y, -\xi, -\eta; t),$$

we get ultimately

$$G(x, y, \xi, \eta, t) = \frac{4}{\pi^2} \int_0^\infty \int_0^\infty \exp[-k(\alpha^2 + \beta^2)t] \sin \alpha x \sin \beta y \sin \alpha \xi \sin \beta \eta \, d\alpha \, d\beta.$$

Case 2. $\frac{\partial G}{\partial x} = 0$ for $x=0$,

$$\frac{\partial G}{\partial y} = 0 \text{ for } y=0.$$

The expression $G = u_1 - u_2 - u_3 + u_4$ evidently satisfies the given boundary conditions. This expression ultimately becomes

$$G = \frac{4}{\pi^2} \int_0^\infty \int_0^\infty \exp[-k(\alpha^2 + \beta^2)t] \cos \alpha x \cos \beta y \cos \alpha \xi \cos \beta \eta \, d\alpha \, d\beta.$$

Case 3. $G=0$ for $x=0$,

$$\frac{\partial G}{\partial y} = 0 \text{ for } y=0.$$

Clearly the expression $G = u_1 + u_2 - u_3 - u_4$ satisfies the given boundary conditions, and may be ultimately written in the form

$$G = \frac{4}{\pi^2} \int_0^\infty \int_0^\infty \exp[-k(\alpha^2 + \beta^2)t] \sin \alpha x \cos \beta y \sin \alpha \xi \cos \beta \eta \, d\alpha \, d\beta.$$

Case 4. $\frac{\partial G}{\partial x} = 0$ for $x=0$.

Interchanging x and y in the preceding formula, we obtain

$$G = \frac{4}{\pi^2} \int_0^\infty \int_0^\infty \exp[-k(\alpha^2 + \beta^2)t] \cos \alpha x \sin \beta y \cos \alpha \xi \sin \beta \eta \, d\alpha \, d\beta.$$

$$\text{Domain B} \begin{cases} 0 < x < \infty, \\ 0 < y < \infty, \\ -\infty < z < \infty. \end{cases}$$

It can be readily seen that for each of the four types of boundary conditions considered in the case of domain A, the corresponding solution may be obtained by replacing the double integral by a triple integral and adding the factor $\frac{2}{\pi} \cos \gamma (z - \zeta)$ to the integrand. Thus in the case of the boundary conditions $G=0$ for $x=0$ and $y=0$ the solution is

$$G = \frac{8}{\pi^3} \int_0^\infty \int_0^\infty \int_0^\infty \exp[-k(\alpha^2 + \beta^2 + \gamma^2)t] \\ \times \sin \alpha x \sin \beta y \sin \alpha \xi \sin \beta \eta \cos \gamma(z - \zeta) \, d\alpha \, d\beta \, d\gamma.$$

The solutions corresponding to the vanishing of the temperature gradient on

- (1) the plane $y=0$,
- (2) the plane $z=0$,
- (3) the plane $y=0$ and $z=0$,
- etc.,

are obtained in a similar fashion.

$$\text{Domain C} \begin{cases} 0 < x < \infty, \\ 0 < y < \infty, \\ 0 < z < \infty. \end{cases}$$

Case 1. $G=0$ for $x=0$,
 $y=0$,
 $z=0$.

In this case we start with

$$u_1 = u_1(x, y, z, \xi, \eta, \zeta; t) = \frac{1}{8[\pi kt]^{3/2}} \exp \left[-\frac{(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2}{4k(t)} \right].$$

With u_1 we associate the functions

$$u_2 = -u_1(x, y, z, \xi, \eta, -\zeta; t), \\ u_3 = -u_1(x, y, z, \xi, -\eta, \zeta; t),$$

$$\begin{aligned}
u_4 &= u_1(x, y, z, \xi, -\eta, -\zeta; t), \\
u_5 &= -u_1(x, y, z, -\xi, \eta, \zeta; t), \\
u_6 &= u_1(x, y, z, -\xi, \eta, -\zeta; t), \\
u_7 &= u_1(x, y, z, -\xi, -\eta, \zeta; t), \\
u_8 &= -u_1(x, y, z, -\xi, -\eta, -\zeta; t).
\end{aligned}$$

The function $G = u_1 + u_2 + \dots + u_8$ satisfies the given boundary conditions, and we ultimately get

$$\begin{aligned}
G = \frac{8}{\pi^3} \int_0^\infty \int_0^\infty \int_0^\infty \exp[-k(\alpha^2 + \beta^2 + \gamma^2)t] \\
\times \sin \alpha x \sin \beta y \sin \gamma z \sin \alpha \xi \sin \beta \eta \sin \gamma \zeta \cdot d\alpha d\beta d\gamma.
\end{aligned}$$

It can be readily seen that if $\sin \alpha x \sin \alpha \xi$ is replaced by $\cos \alpha x \cos \alpha \xi$ we obtain the solution corresponding to the boundary conditions

$$\frac{\partial G}{\partial x} = 0 \text{ for } x=0 \text{ and } G=0 \text{ for } y=0 \text{ and for } z=0.$$

Yeshiva College,
New York, N.Y.

LVII. Notices respecting New Books.

Fourier Series. By G. H. HARDY and W. W. ROGOSINSKI. [Pp. viii. + 100.] (Cambridge University Press, 1944.) Price 8s. 6d.

FOURIER was led to trigonometrical series by a physical problem; but in this little book the relationship of Fourier Series to physical or other practical problems has vanished, and we are confronted with pure mathematics. The authors tell us in the preface that they "have not written for physicists or for beginners." Few books contain so much in 100 pages. This is achieved by avoiding repetitions and by the systematic use of abbreviations. A Fourier series is a (F.s.); Fourier constants are (F.c.), and so on, and one feels that it makes the book more laborious to study than it might have been without such rigorous abbreviation and with more repetition. The work begins where many of us have left off. The coefficients with the familiar integral form do, indeed, appear; but Fourier and Dirichlet, as the authors themselves say, "are hardly mentioned." It is assumed that the reader will be acquainted with a great deal in advance, including Lebesgue's theory of measure and integration. And here a word or two may be written for the physicist, with mathematical instincts, who may feel a bit depressed by the prefatory reference to himself, quoted above. The appearance of such terms as *measure*, *mean*, *space*, *orthogonal system*, *metric*, and still others, including the *Gibbs phenomenon*, shows how intimate is the correlation between even the modern theory of Fourier Series and physics.

The book deals very ably with a part of pure mathematics. It is not concerned with any sort of application outside pure mathematics. Such works also are of value.

W. W.

[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]

LVIII. *A Relativistic Theory of Temperature Radiation.*

By HERBERT DINGLE, Professor of Natural Philosophy, Imperial College of Science and Technology *.

[Received July 19, 1944.]

ABSTRACT.

The phenomenon of temperature radiation is described in relativistic form—i. e. in terms not involving the conception of interchange of radiation between bodies at the same temperature, since such radiation, like absolute motion, is unobservable. Temperature is measured, relatively to an arbitrary zero, by the time rate of change of a quantity η , which is closely related to the entropy change of the radiating body, and both η and time are measured by strictly defined thermal processes. Transformation equations, analogous to the Lorentz equations in mechanical relativity, are calculated, for use when the arbitrary zero of temperature is changed, and a thermal "interval," invariant under such transformations, is derived. The process of temperature radiation is thus given a geometrical interpretation, and takes a form in which the powerful tensor calculus is available for its study. Only the "special" theory is worked out, applicable to bodies radiating at constant temperature, just as the special theory of mechanical relativity refers only to bodies moving with constant velocity. The physical meaning of the concepts employed, and the field of application of the special theory, are discussed, and the method of generalization is indicated.

The Parallel Relationships of Motion and Temperature.

THERE are two observable properties of matter which depend on the whole universe: they are motion and temperature. Introduce a new piece of matter into the universe, and the motion (through gravitation) and temperature (through radiation) of every other piece of matter are changed. Similarly, if any existing piece of matter be moved, the motion of all other matter is altered; and if any existing piece of matter be heated or cooled, the temperature of all other matter is altered. These fundamental facts are expressed in current physical theory by the two universal laws of motion and gravitation on one hand, and the second law of thermodynamics on the other.

There is apparently a distinction between motion and temperature in that the former necessarily involves a reference to time (motion is regarded

* Communicated by the Author.

as change of position with time), while the latter is essentially static—a passive condition of a body indicated by the equilibrium reading of a thermometer. In many philosophical discussions, for example, motion is taken as almost synonymous with change, while constant temperature is regarded as a necessary attribute of an eternally unchanging world. This distinction, however, is quite arbitrary; it resides not in the phenomena themselves but in the terms in which we choose to measure them. It is customary to represent motion by velocity, defined for measurement purposes as dx/dt , but velocity might equally well be defined by the Doppler effect and measured as proportional to $d\lambda/\lambda$. In that case time would not be involved; the velocity of a body would be indicated by the *instantaneous* position of a spectrum line and not by a *change* of position between two successive instants of time. Similarly, instead of the customary measurement of the temperature of a body by the instantaneous reading of a thermometer, we might adopt a measurement in terms of the rate of radiation of energy, viz. dE/dt . In that case temperature would denote a process, and so have reference to time.

It will be convenient to set out the alternative methods of measurement side by side, taking, for simplicity of expression, $\frac{d(\lambda^2)}{\lambda^2}$ instead of $\frac{d\lambda}{\lambda}$ for the Doppler measure of velocity. This is done in Table I.

TABLE I.

<i>Motion.</i>		<i>Temperature.</i>
Kinematic	$v = \frac{dx}{dt}$.	$\psi \equiv \frac{dE}{dt} = a(T + 1/\alpha)^4 - b(T' + 1/\alpha)^4$.
Static	$V \equiv \frac{d(\lambda^2)}{\lambda^2} = \frac{2\frac{v}{c}}{1 - \frac{v}{c}}$.	$T = \frac{P_T}{\alpha P_0} - \frac{1}{\alpha}$.

A word of explanation of the Table is desirable. According to the accepted theory of the Doppler effect, the wave-length, λ' , of a spectrum line from a source moving with velocity $v \left(= \frac{dx}{dt} \right)$ away from an observer is given by

$$\lambda' = \lambda \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}, \dots \dots \dots (1)$$

where λ is the wave-length of the same line from a stationary source, and c is the velocity of light. It follows immediately that if we adopt $d(\lambda^2)/\lambda^2$ as a measure of velocity, and call it V , the relation between V

and v is that given in the Table. On the temperature side, the "static" measure taken is that of the perfect-gas constant-pressure thermometer; α is the coefficient of pressure increase of a perfect gas, and P_T and P_0 the pressures indicated by the thermometer at $T^\circ\text{C.}$ and 0°C. respectively. For the "kinematic" measure, the Stefan law is adopted, a and b being constants and T' the appropriate average temperature of the surroundings.

Before proceeding it will be advisable to note one point which is often misunderstood. The existence of a limiting value for velocity or temperature has nothing to do with those properties themselves, but depends solely on the method of measurement adopted. The limits of $v\left(=\frac{dx}{dt}\right)$

are $\pm c$, the velocity of light, but if we measure velocity by V , the limits are ∞ and -1 ; *i. e.* a velocity of recession may be infinite but a velocity of approach cannot exceed the velocity of light. The reason is, of course, that a wave-length can be increased without limit, but cannot be decreased below zero.

On the temperature side, since P_T may be infinite but cannot be less than zero, T lies between ∞ and $-\frac{1}{\alpha}$ —the familiar "absolute zero" of temperature. ψ , however, may be anything between $+\infty$ and $-\infty$, the former corresponding to a body with $T=\infty$ and $T'=-\frac{1}{\alpha}$, and the latter to a body with $T=-\frac{1}{\alpha}$ and $T'=\infty$. If, for any reason, it were desirable to have an "isotropic" temperature scale with finite limits, we could choose

$$\phi = k \frac{P_T - P_0}{P_T + P_0}, \quad \dots \dots \dots (2)$$

where k is any constant. This perfectly unambiguous, directly applicable, and altogether satisfactory scale would give temperatures lying between $+k$ for $P_T=\infty$ and $-k$ for $P_T=0$. The limiting velocity, c , and the absolute zero of temperature are therefore not signs of limitation on the behaviour of matter, but characteristics of our methods of measurement.

Relativistic and Non-relativistic Systems of Measurement.

In Table I., the correspondence between the measurements of motion and temperature is marred by one difference—namely, that the former are relativistic and the latter "absolutist." Before Einstein's theory was accepted the measurements of velocity would have been given as

$$v = v_0 + \frac{dx}{dt} \quad \dots \dots \dots (3)$$

$$\text{and} \quad \bar{V} \left(\text{defined as } \frac{d\lambda}{\lambda} \right) = \bar{V}_0 + \frac{v}{c}, \quad \dots \dots \dots (4)$$

where v_0 and \bar{V}_0 represent the (unknown) absolute velocity of the arbitrarily chosen standard of rest. It is now recognized that v_0 and

\bar{V}_0 are meaningless and should be excluded from our formulæ. No such recognition, however, has been extended to temperature measurement, and the term $\frac{1}{\alpha}$ —the analogue of v_0 —appears as the “absolute” temperature of the arbitrarily chosen zero. In view of the progress made when the relativity of motion was acknowledged, the question arises whether like progress might not be made by acknowledging the relativity of temperature, so far, at least, as the phenomenon of radiation is concerned.

Objections immediately suggest themselves, but do not seem to be final. The most obvious is that we have discovered an absolute temperature but have never discovered an absolute velocity. This difference, however, is more apparent than real. In so far as the “discovery” of an absolute temperature depends on the existence of a limiting zero, whose value can be assigned on the arbitrary Centigrade scale, we have seen that this is a characteristic of our system of measurement only. What we have discovered is that we can measure temperature from a limit which can be evaluated as about -273°C . But we have discovered this also with regard to motion. We have here a limit which can be evaluated as about $3 \cdot 10^{10} \text{ cm./sec}$. The relativity of motion does not depend on the existence or otherwise of a particular absolute velocity. What it depends on is the postulate that we can take the velocity of any body we wish as a zero velocity, and that our expressions of laws of nature in terms of our resulting measurements shall be the same as those arrived at when the velocity of any other body is chosen as zero. Whether this is so in every respect with regard to temperature is a question which is outside the scope of the present discussion, for we are concerned here with temperature only in relation to radiation. The question we have to consider is this: Do the reasons which have led to the adoption of a relativity theory of motion apply equally strongly to the theory of radiation? It is maintained here that they do, and the beginnings of a relativity theory of radiation have been attempted.

Reasons for Attempting a Relativistic Theory of Radiation.

The fundamental fact is that just as the motion (if any) of two bodies having no relative motion has no effect on their mutual relations, so the radiation (if any) of two bodies having no temperature difference has no effect on their mutual relations. In pre-relativity theory, two bodies whose relative velocity was v were regarded as having velocities v_0 and $v_0 + v$, the relative velocity being the difference between the two. If the bodies were relatively at rest, they each had a velocity v_0 . This was discarded because v_0 had no physical effects. In the phenomenon of radiation we find that two bodies at different temperatures radiate to one another while two bodies at the same temperature undergo no change. If θ is the temperature of one with respect to the other, then we say that they have absolute temperatures θ_0 and $\theta_0 + \theta$, and, further,

that each radiates to the other a quantity of energy dependent on its temperature, the *observable* radiation being only the difference between the two quantities. When the bodies are at the same temperature there is no observable radiation between them. Nevertheless, we say that each radiates at a definite, absolute rate, although the radiation is unobservable. There appears to be no more justification for this than there was for speaking of absolute velocities.

The analogy may be carried a step further, from the special into the general theory of relativity, although in this paper we shall not pursue this side of the matter, being concerned only with the special theory. In pre-relativity mechanics a body was regarded as having two independent properties—inertial mass and gravitational mass—which happened always to be equal to one another. In radiation theory similarly a body is regarded as having two independent properties—emissive power and absorptive power—which happen always to be equal to one another. General relativity removed the accidental character of the identity between inertial and gravitational mass by describing motion as a relation between bodies and not as a property of each of them. It seems desirable, therefore, to attempt to remove the accidental character of the identity between emissive and absorptive powers by describing temperature radiation as a relation between bodies and not as a property of each of them.

The Process by which Mechanics became Relativistic.

We seek to re-express the law of radiation in relativistic terms. We shall consider constant temperatures only, just as in special mechanical relativity we consider constant velocities only. The theory will therefore not apply exactly to any actual situation, although it may provide a close approximation to actuality in certain circumstances. Just as the special mechanical theory is practically valid in an isolated system whose mass is small, so the special radiation theory will be practically valid in an isolated system whose heat capacity is very large, so that radiation does not appreciably change its temperature. A normal star is an example.

We shall follow the procedure already found successful in kinematics, which may be briefly summarized as follows:—We define motion in terms of the concepts of space and time; *i. e.*, we adopt the definition $v = \frac{dx}{dt}$ for velocity. We measure space by choosing a unit measuring rod, and we then define equal intervals of time as those taken by a body unacted upon by forces (or, if we wish, by a beam of light: Einstein's assumption of the "constancy" of the velocity of light was simply an assumption that light would give the same time-scale as a body unacted upon by forces) to cover equal spaces. Having thus defined the measurement of space and time, we measure velocity by the value of dx/dt . Motion having thus been represented in metrical terms, we define other quantities—mass, force, etc.—in terms of the same concepts of space, time

and velocity, and generalize our observations of the behaviour of bodies in expressions called laws of nature, which are formulæ relating our measurements together.

Suppose, now, we wish to regard the whole mechanical system we are examining as having a velocity v with respect to some arbitrary external standard. We must then add v to all the velocities, $u_1, u_2 \dots$, in the system. But the theory of relativity requires that v must cancel out when the laws of nature are arrived at, since those laws must be the same whatever value we care to assign to v . The condition that this shall be so is that we shall define the length of a body as $l\sqrt{1-v^2/c^2}$, where l is the number of times a relatively stationary standard scale is contained in the body, and v is the velocity the body is assumed to have. This results in the well-known transformation equations (we give one space dimension only, for simplicity, and write \bar{t} for time, since t will be given another meaning later):—

$$\left. \begin{aligned} dx' &= \frac{dx - v d\bar{t}}{\sqrt{1 - \frac{v^2}{c^2}}} \\ d\bar{t}' &= \frac{d\bar{t} - \frac{v}{c^2} dx}{\sqrt{1 - \frac{v^2}{c^2}}} \\ u' &= \frac{u - v}{1 - \frac{uv}{c^2}} \end{aligned} \right\} \dots \dots \dots (5)$$

with corresponding equations for mass, force, etc. It then follows that

$$c^2 d\bar{t}'^2 - dx'^2 = c^2 d\bar{t}^2 - dx^2 \equiv ds^2, \quad \dots \dots \dots (6)$$

the so-called “interval.”

It will be noticed that the essence of this system of kinematics is the measurement of motion, or velocity, in terms involving time. Had it been measured in terms of the Doppler effect—of V instead of v —a system of kinematics satisfying the relativity principle might have been constructed, but it would have been a fundamentally different system. It would have been at bottom, of course, an equivalent system, but we should have had to go to fundamentals to find the correspondence. In the system actually adopted, we choose a concept of space, then measure time *in terms of the measure of space*, and so make possible the construction of the invariant quantity, ds , which is a combination of the measurement of space and the related measurement of time. Time is thus employed as an auxiliary, so to speak, to amplify our concept of space into a “space-time,” which makes possible a description of observable motions without the intrusion of unobservable absolute motions.

Let us now proceed in a similar way in dealing with radiation. We

define radiation in terms of the concepts of energy and time (we shall, in actual fact, substitute for energy a quantity more nearly related to entropy, but in setting out the principles of the theory it seems preferable to speak in terms of more familiar and quite legitimate, though mathematically less convenient, concepts); i. e. we adopt a definition $\psi = \frac{dE}{dt}$ for temperature. We measure E by a suitably defined instrument, and we then define equal intervals of time as those taken by a specified body to radiate equal amounts of energy. Having thus defined the measurement of energy and time, we measure temperature by the value of $\frac{dE}{dt}$.

Suppose, now, we wish to regard the whole radiating system we are examining as having a temperature ψ_0 with respect to some arbitrary external standard. We must then add ψ_0 to all the temperatures in the system. But our theory of relativity requires that ψ_0 must cancel out when the laws of nature are arrived at, since those laws must be the same whatever value we care to assign to ψ_0 . We must therefore determine the transformation equations that will ensure this, and so reach the thermal equivalent of the "interval," which will be invariant with respect to such transformations.

In this process we have one great advantage compared with those who originally derived the transformation equations for motion, in that we are familiar with observations relatively much nearer to the absolute zero of temperature than ordinary velocities are to the velocity of light. The transformation equations can therefore be obtained directly from well-known experimental results, instead of being arrived at indirectly and capable of test only in very exceptional cases.

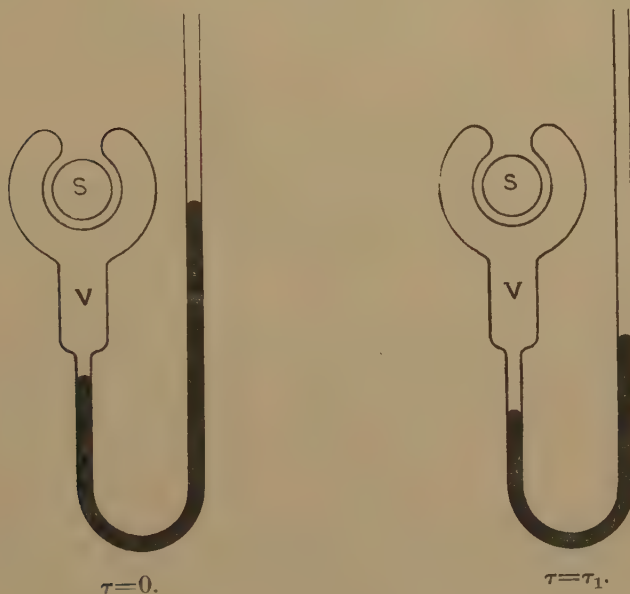
Measurement of Radiation and Time.

(a) *Radiation.*—We choose as our instrument for measuring radiation a vessel containing one gram-molecule of a perfect gas at constant temperature, with a manometer attached to indicate its pressure. (Strictly speaking, of course, we cannot yet say what we mean by constant temperature; and, indeed, when the theory is generalized it will not be necessary to do so, for any temperature may be regarded as constant, just as in general relativity any motion may be regarded as rest. But for a special theory we have to choose some practical criterion, just as, in the special theory of relativity of motion, we choose some practical criterion of the constancy of the velocity v in the transformation formulæ. It will suffice for present purposes to stipulate that an ordinary thermometer, immersed in the perfect gas, shall keep a constant reading.) If, now, a body at the same temperature is brought into the neighbourhood of this instrument, the pressure of the gas in the instrument will not change; but if a body of higher or lower temperature is brought near, the pressure will fall or rise, respectively, as the gas receives heat

is not perfectly rigid or maintained always at absolutely the same temperature.

This instrument satisfies the essential condition that it can be defined independently of temperature. It is necessary that the temperature shall not change, but it may have any value, and we may compare the readings of two such instruments at different temperatures. In fig. 1, the zero readings of two such instruments are shown, the temperatures being 0 and τ_1 . We shall later derive transformation equations connecting the readings of these two instruments. The temperature of either could, of course, have been chosen as zero, just as we may choose any desired zero of velocity.

Fig. 1.



(b) *Time*.—A kinematic clock consists in principle of a specified kind of body moving over a space-measuring scale. The thermal clock will therefore consist of a specified kind of body radiating to an η -measuring instrument. Diagrammatically it is identical with fig. 1, the “specified kind of body” being placed in the sphere S. We choose as such a body one gm. mol. of a perfect gas having the same volume as the initial volume of the gas in V but a slightly different (say higher) pressure, $p_1 + dp_1$. All this can be realized without measurement of temperature, but we know from experience that the gas in S will then be at a slightly higher temperature than that in V, given in ordinary terms by

$$\theta_0 + d\theta_0 = \theta_0 \frac{p_1 + dp_1}{p_1} \quad \dots \quad (8)$$

We shall then have continuous radiation from S to V, the temperature of S being maintained constant by supplying heat, and that of V being

maintained constant by expansion or contraction of the gas. The unit of time will naturally be the time required for V to acquire one unit of η . Its value can be made what we please by a suitable choice of dp_1 . We will not yet fix it definitely, but we must impose some restriction on it in order to ensure that the clock shall have a definite specification whatever the temperature of our instrument. We will fix dp_1 as a constant fraction, ν , of p_1 , and ν may be chosen to suit our convenience.

This instrument, like that for measuring η , can be used at any temperature, for whatever the temperature may be, the requirement that the gas in S shall have the same volume but a higher pressure than the initial values of that in V ensures that it shall radiate continuously. Winding up the clock corresponds to supplying the requisite amount of heat to S. We shall always, of course, keep the gas in V at the same temperature in both the η -measuring instrument and the clock, just as the space-measuring instrument and clock have the same velocity in mechanical measurements.

The Equations of Transformation.

We consider two pairs of instruments at different temperatures. We call the temperature of one pair zero, and that of the other τ_1 , and we seek the relations between η , t and $\tau \left(= \frac{d\eta}{dt} \right)$, measured by the first pair, and η' , t' and $\tau' \left(= \frac{d\eta'}{dt'} \right)$ measured by the second pair, when a given body is examined by them. To find these relations, we make use of known facts, usually expressed in terms of the customary temperature scale; for these facts are true, however they may be expressed. The ordinary expressions, of course, will be absent from the final transformation equations.

Consider first the transformation for the η -measuring instruments. Suppose the (ordinary) absolute temperature of the instrument at the arbitrary zero is θ_0 , and that of the body to be tested is θ . Then, during an interval of time measured by an ordinary mechanical clock as $d\bar{t}$, the heat radiated from the body to the instrument is

$$dH = a(\theta^4 - \theta_0^4)d\bar{t}, \quad \dots \dots \dots (9)$$

the outer surface of S and the adjacent surface of V being regarded as identical, so that the same constant a applies to each. Hence

$$a(\theta^4 - \theta_0^4)/d\bar{t} = -R\theta_0 \log \frac{p_2}{p_1}, \quad \dots \dots \dots (10)$$

so that

$$d\eta \equiv \beta \log \frac{p_2}{p_1} = -\frac{\beta a}{R\theta_0}(\theta^4 - \theta_0^4)d\bar{t}. \quad \dots \dots \dots (11)$$

Now consider the value of the interval of time, $d\bar{t}$, in terms of the readings of the thermal clock. During this interval the heat radiated by S to V in the clock will be

$$dH = a[\theta_0^4(1 + \nu)^4 - \theta_0^4]d\bar{t}, \quad \dots \dots \dots (12)$$

since, from (8), $d\theta_0 = \theta_0 \frac{dp_1}{p_1} = \nu \theta_0$. Hence in this time the amount of η received by V will be

$$-\frac{\beta a \theta_0^4}{R \theta_0} [(1+\nu)^4 - 1] d\bar{t}, \quad \dots \quad (13)$$

and since the measure of thermal time is defined as proportional to the amount of η received, we may represent this by dt units of thermal time, so that

$$dt = -\frac{\beta a \theta_0^3}{R} [(1+\nu)^4 - 1] d\bar{t}, \quad \dots \quad (14)$$

an equation which gives the general relation between mechanical and thermal time.

Substituting for $d\bar{t}$ in (11) we therefore obtain

$$d\eta = \left(\frac{\theta^4}{\theta_0^4} - 1 \right) \frac{1}{[(1+\nu)^4 - 1]} dt. \quad \dots \quad (15)$$

Hence the temperature of the body under consideration is

$$\tau \equiv \frac{d\eta}{dt} = \left(\frac{\theta^4}{\theta_0^4} - 1 \right) \frac{1}{[(1+\nu)^4 - 1]} = \zeta \left(\frac{\theta^4}{\theta_0^4} - 1 \right), \quad \dots \quad (16)$$

where

$$\zeta = [(1+\nu)^4 - 1]^{-1}. \quad \dots \quad (17)$$

Equations (11), (14) and (16) give the thermal quantities η , t and τ in terms of the ordinary thermal quantities. With the substitution (17), they become

$$\left. \begin{aligned} d\eta &= -\frac{\beta a}{R \theta_0} (\theta^4 - \theta_0^4) d\bar{t} \\ dt &= -\frac{\beta a \theta_0^3}{R \zeta} d\bar{t} \\ \tau &= \zeta \left(\frac{\theta^4}{\theta_0^4} - 1 \right) \end{aligned} \right\} \quad \dots \quad (18)$$

We can now at once obtain the transformation equations. Suppose the temperature, τ_1 , of the second instrument corresponds to absolute temperature θ_1 . Then, for this instrument, equations (18) become *

* It is to be noted here that by choosing the same value of $d\bar{t}$ for the two sets of instruments we are following the lead given by mechanical relativity in regarding the interval between two events *in the experience of the observer* as absolute. The duration of time in which the body radiates to each of the two η -measuring instruments is an interval of *proper* time: we may, if we wish, make it actually the same interval of time for both sets by having two identical bodies and placing one of them in each instrument. The values of $d\eta$, dt , $d\eta'$, dt' will then correspond to a single duration in the experience of the observer.

$$\left. \begin{aligned} d\eta' &= -\frac{\beta a}{R\theta_1}(\theta^4 - \theta_1^4) d\tilde{t} \\ dt' &= -\frac{\beta a \theta_1^3}{R\zeta} d\tilde{t} \\ \tau' &= \zeta \left(\frac{\theta^4}{\theta_1^4} - 1 \right) \end{aligned} \right\} \dots \dots \dots (19)$$

Further, we have, from (18)

$$\tau_1 = \zeta \left(\frac{\theta_1^4}{\theta_0^4} - 1 \right) \dots \dots \dots (20)$$

From (18), (19) and (20) we have to eliminate θ , θ_0 and θ_1 . The result is easily found to be

$$\left. \begin{aligned} d\eta' &= \frac{d\eta - \tau_1 dt}{(1 + \tau_1/\zeta)^{1/4}} \\ dt' &= (1 + \tau_1/\zeta)^{3/4} dt \\ \tau' &= \frac{\tau - \tau_1}{1 + \tau_1/\zeta} \end{aligned} \right\} \dots \dots \dots (21)$$

These are the required transformation equations. They correspond, in the theory of radiation, to the Lorentz transformation equations in the theory of motion.

Unlike the Lorentz equations, however, these equations, when solved for undashed in terms of dashed quantities, do not yield the same formulæ with a change of sign for τ_1 . This follows from the “non-isotropic” scale of measurement adopted. We obtain

$$\left. \begin{aligned} d\eta &= \frac{d\eta'(1 + \tau_1/\zeta) + \tau_1 dt'}{(1 + \tau_1/\zeta)^{3/4}} \\ dt &= \frac{dt'}{(1 + \tau_1/\zeta)^{3/4}} \\ \tau &= \tau' + \tau_1 + \frac{\tau' \tau_1}{\zeta} \end{aligned} \right\} \dots \dots \dots (22)$$

Incidentally, it may be remarked that these transformations form a group.

The τ temperature scale has no preferential zero value. Just as we may assume any standard of rest in mechanics, so we may assume any temperature to be zero here, and the equations will hold good. The temperature $-\zeta$ —corresponding to the “absolute zero”—has the same value whatever zero we choose, for if we put $\tau = -\zeta$ in (21) we obtain $\tau' = -\zeta$. This is analogous to the invariance of the velocity of light in all kinematic co-ordinate systems.

The Invariant Thermal Interval.

Eliminating τ_1 from the equations of (21) we obtain

$$d\sigma^4 \equiv dt(dt + d\eta/\zeta)^3 = dt'(dt' + d\eta'/\zeta)^3 \dots \dots \dots (23)$$

$d\sigma$ thus corresponds to ds in mechanical relativity: it is invariant with respect to the temperature of the measuring instruments, just as ds is invariant with respect to the velocity of the measuring instruments. It may easily be verified that (21) is the most general linear transformation that preserves the invariance of $d\sigma$.

If we give to thermal relativity the same type of geometrical interpretation that is given to mechanical relativity, we must regard $d\sigma$ as the line element of a "space" whose metric is given by an expression of the 4th degree in $d\eta$ and dt . The null geodesic, given by $d\sigma=0$, then corresponds to $dt=0$ or $d\eta/dt=-\zeta$. In mechanical relativity, $ds=0$ gives $dx/dt=\pm c$. The difference arises from the fact that the limiting values of v are $\pm c$, whereas those of τ are $+\infty$ and $-\zeta$, as we may see from the last equation of (18) by giving θ its extreme values of ∞ and 0 . This, as we have already pointed out, is a characteristic of our scale of measurement, and does not denote some fundamental physical difference between the relations of velocity and those of temperature. $dt=0$ may, of course, also be written $d\eta/dt=\infty$, so that the null geodesic corresponds to the limits of temperature, $+\infty$ and $-\zeta$.

Mathematical Aspect of the Theory.

In the foregoing sections the theory has been presented as a development from a general principle, which is held to be applicable throughout science. Many physicists, however, distrust theories originating in this way, feeling that such principles must give a metaphysical tinge to everything deduced from them—although, in fact, the particular principle involved here is simply that science should be independent of metaphysics. It will not be superfluous, therefore, to restate the theory in purely mathematical terms, from which it will be seen that it expresses only facts already known and accepted, but in terms of unusual functions of measured quantities. The reason for choosing such functions will, of course, be obscure, because they belong to the aspect of the theory which is here deliberately veiled, but it will be made clear that the mathematical theory involves no new assumption, presupposes no kinds of measurement which cannot be—in fact, which have not been—carried out, and contains no circular argument.

We wish to describe the process of radiation from a black body maintained at a constant absolute temperature θ , keeping strictly within the compass of familiar established ideas and definitions. If A is the area of the body and ϕ is Stefan's constant, the energy radiated in time \bar{dt} is

$$dH = A\phi\theta^4\bar{dt}.$$

ϕ is a known quantity, so this can actually be evaluated. Suppose this energy is absorbed isothermally by one gm. mol. of a perfect gas in a black-body enclosure of surface area A at temperature θ_0 , thus changing the pressure of the gas from p_1 to p_2 . Then we know that since, during the process, the gas will receive energy $A\phi\theta^4\bar{dt}$ and radiate energy

$A\phi\theta_0^4 d\bar{t}$, the difference between these quantities, viz. $A\phi(\theta^4 - \theta_0^4) d\bar{t}$, will be equal to $-R\theta_0 \log \frac{p_2}{p_1}$; and so, if β is an arbitrarily chosen constant, we shall have

$$\beta \log \frac{p_2}{p_1} = -\frac{\beta A\phi}{R\theta_0} (\theta^4 - \theta_0^4) d\bar{t}.$$

Call this quantity $d\eta$. We then have equation (11), with $a=A\phi$

Next, imagine two separate gm. mols. of a perfect gas, initially of equal volume, each enclosed in a vessel having a black-body surface of area A , and let them be arranged so that all the radiation from one is absorbed by the other. Let the pressure and temperature in one be p_3 and θ_0 , respectively, and let the pressure in the other be $p_3(1+\nu)$, where ν is an arbitrarily chosen constant. We then know that the temperature in the second will be $\theta_0(1+\nu)$. The balance of radiation received by the first from the second in time $d\bar{t}$ will therefore be $A\phi[\theta_0^4(1+\nu)^4 - \theta_0^4] d\bar{t}$, heat being supplied, as may be necessary, to the second to keep its temperature constant at $\theta_0(1+\nu)$, and the pressure in the first changing from p_3 to p_4 so that its temperature remains constant at θ_0 . We then know, from the accepted properties of perfect gases, that

$$A\phi[\theta_0^4(1+\nu)^4 - \theta_0^4] d\bar{t} = -R\theta_0 \log \frac{p_4}{p_3},$$

so that

$$\beta \log \frac{p_4}{p_3} = -\frac{\beta A\phi\theta_0^3}{R} [(1+\nu)^4 - 1] d\bar{t},$$

and the right-hand side of this equation is obviously calculable.

Call this quantity $d\bar{t}$. We then have equation (14) with $a=A\phi$.

Finally, divide $d\eta$ by $d\bar{t}$, obtaining $d\eta/d\bar{t}$. *Call this quantity τ .* We then have equation (16).

The new conceptions of the theory are thus completely expressed by the choice of three symbols, named in the italicized sentences above. According to the ordinary expression of the law of radiation we have

$$dH = a\theta^4 d\bar{t}.$$

We propose to re-express this as

$$d\eta = \tau d\bar{t}.$$

It is obvious, then, that what we are proposing is merely a change of variables which can actually be made from knowledge which already exists.

It may be asked what advantages follow from this re-expression. The answer is that they are two. First, the new law does not imply, as does the old, that radiation exists which is essentially undetectable. We cannot observe dH , but only dH reduced by the radiation assumed to be received by the body from its surroundings, and this balance becomes zero when the surroundings are at the temperature of the body. Nevertheless, dH is assumed to "exist." $d\eta$, on the other

hand, automatically vanishes when the temperature of the instrument is that of the radiating body. This is an advantage of principle; it may be compared to the advantage which the special theory of relativity has over the material ether theory supplemented by the hypotheses of the Fitzgerald contraction and time dilatation. The second advantage will be rated more highly by those to whom achievement in science is more important than understanding of what is achieved. It is that in terms of the new law we may deduce an invariant quantity, $d\sigma$, characteristic of the radiation of the body and expressible in terms of the "co-ordinate differences," $d\eta$ and dt . A geometrical interpretation of the process of radiation is thus made possible, and the powerful tensor calculus is made available for its further study. For, by repeating the above process with a different value, θ'_0 , for θ_0 , we obtain quantities $d\eta'$, dt' and τ' , and the equations of transformation, obtained by simple algebra, show that $dt(dt+d\eta/\zeta)^3 = dt'(dt'+d\eta'/\zeta)^3$.

The quantities $d\eta$, dt , τ and $d\sigma$ are here introduced simply as mathematical functions. In the next section we shall consider their significance in relation to current physical conceptions, but it might serve as a useful introduction to that section to explain why they have been given certain names. $d\eta$ and $d\sigma$ have been left undenominated, but dt has been called "thermal time" and τ "temperature." From the purely mathematical point of view taken in this section the names need some justification.

I have called dt thermal time because, as equation (14) shows, it is directly proportional to an interval of time as ordinarily defined, and therefore is numerically identical therewith when the appropriate units are chosen. Since the principle of determination, however, depends on a thermal instead of a kinematical process, I have distinguished dt from dt by the adjective "thermal." Those, therefore, who feel that the word "time" has already been earmarked for the results of measurements by the established kinematical process will not be in danger of confusion.

The choice of the term "temperature" for τ , however, involves slightly different considerations. τ is not proportional to θ , but it does measure the general property which we call temperature, because all bodies which have the same θ have the same τ , and of any two bodies, that which has the higher θ has the higher τ . It would seem appropriate, therefore, to call τ the "radiation temperature," but I have omitted the adjective in this case because, unlike time, temperature already has several alternative systems of measurement in common use, and they are distinguished not by adjectives but by being called different "scales" of temperature. Thus, we have the Kelvin scale, the mercury-in-glass scale, the platinum resistance scale, etc., all of which give different values for the same "temperature"; but we do not call those values the "work temperature," "expansion temperature," "resistance temperature," but "temperature on the work scale," "temperature on the mercury-in-glass scale," "temperature on the platinum resistance scale." I have therefore allowed τ to represent simply "temperature on the

τ -scale," which is particularly appropriate to the study of radiation as the Kelvin scale, for example, is particularly appropriate to the study of the transformation of heat into work.

It would be a gross error, however, to interpret the present theory, which is simply an expression of the process of radiation in a form in which it is invariant to change of zero on the τ -scale, as implying that all physical phenomena are invariant to a change in the zero of temperature. The theory in the first instance concerns radiation only, because it is radiation, and not necessarily all thermal phenomena, that is detectable only when there are temperature differences. In a later section I discuss other phenomena to which the theory might be extended, but I do not suggest that its application is universal, as is that of the relativity theory of motion. The general criterion is this. Consider a constant temperature enclosure. The theory should be capable of extension to all thermal phenomena which do not enable a person in the enclosure to say what the temperature is, but not to phenomena which may do so. I am not prepared to say what those phenomena are.

Interpretation of Thermal Relativity in Terms of Current Physical Conceptions.

Returning now from the bare mathematical to the physical aspect of the theory, we may easily obtain the equivalents of η , t and σ in terms of the ordinary concepts of thermodynamics. We have a radiating body, and we seek to describe its radiation in terms of the readings of

certain instruments. The first measures $d\eta = \beta \log \frac{p_2}{p_1} = -\beta \frac{dE}{R\theta_0}$, where

dE is the energy received by the instrument during the period under consideration, and θ_0 is the temperature of the instrument. η is thus proportional to the entropy received by the instrument.

No difficulty is associated with dt , for it is simply a period of time; and it may be noted again that the thermal clock gives the same scale as our ordinary mechanical clocks, for since S and V in the clock are kept at constant temperatures, the rate of radiation from one to the other remains constant. The only difference between t and ordinary time \bar{t} is in the value of the unit (and this difference may be removed at any desired temperature by a suitable choice of ν), and in the transformation formula. In any one system of co-ordinates (*i.e.* at any one temperature) an ordinary clock can be rated to read thermal time.

The meaning of $d\sigma$ can be arrived at by expressing dt and $d\eta$ in terms of ordinary concepts, from (18). We obtain

$$d\sigma = \frac{\beta}{R\zeta} \cdot \frac{a\theta^4 d\bar{t}}{\theta} \cdot \dots \dots \dots (24)$$

But $a\theta^4 d\bar{t}$ is, according to the ordinary view, the energy radiated by the body during the time $d\bar{t}$; hence, since the radiation takes place at the

constant temperature θ , $\frac{a\theta^4 \bar{dt}}{\theta}$ is the entropy lost by the body during that time. $d\sigma$ is thus proportional to the *gross* entropy loss of the body, not the net loss when the entropy received from the instrument is taken into account. It is this gross loss that is invariant with respect to the temperature of the measuring instrument.

While this interpretation of the relativity symbols in terms of current concepts may be very useful in facilitating applications, it must be understood that actually the relativity formulæ represent an entirely new outlook, in which the former concepts have no significance. Again the situation is quite analogous to that of mechanical relativity. The mechanical formulæ were formerly looked upon as representing the behaviour of a material body, ploughing its way through a field of ether, and having its internal structure modified ("contracted") by interaction with the ether. Relativity abolished all that. It contemplated no material ether, but only a manifold—"space-time"—in which the body served merely as a landmark to identify particular "points." The space covered, the time taken to cover it, and the invariant interval were all characteristics of space-time, not of the behaviour of the moving body.

In thermal relativity we have the same change. We must no longer think of the "radiating" body as sending out energy into the ether. Its state merely identifies positions in the manifold of " η -time." If we wish to determine the interval between two such positions we must apply our instruments and calculate $d\sigma$ from their readings. If the instruments are at the same temperature as the body, then $d\eta=0$, and the interval is entirely a lapse of time. If the instruments are at a different temperature, then it is partly an interval of η and partly an interval of time. But whatever the temperature of the instruments, $d\sigma$ is the same; it is characteristic not of the behaviour of the body, but of the relation between two points in η -time which the body happens to occupy.

The relation of η to time—of entropy to time—is thus fundamentally of the same character as the relation of space to time. Much irrelevant matter has been written to prove that mechanical relativity has "revealed" an objective identity of space and time, so that one can be changed into the other. The fact is that the "nature" of space and time, whatever may be meant by the phrase, has nothing to do with the case. Space and time are identifiable simply because we have chosen to measure time in terms of space—the space covered by a moving body. If, similarly, we choose to measure time by entropy—the entropy received from a radiating body—then time and entropy become "identified" in the same way. The "nature" of entropy, if the phrase means anything, is again beside the point. Further, if we had chosen energy instead of entropy for the description of radiation, we should have had an "identity" of energy and time. Thermal relativity, even if it

should have no other use, may perform a valuable service by exposing the falseness of the mystical identification of space and time.

Applications.

The preceding re-expression of thermal relations introduces nothing essentially new, for only established results (*e. g.* Stefan's law, the perfect gas equation, etc.) have been used. The new results to which special mechanical relativity led were due to its introduction of a new postulate—namely, that the length of a body (and, in consequence, other physical measurements also) was not a constant but a function of the velocity assigned to the body. The factor $\left(1 - \frac{v^2}{c^2}\right)^{1/2}$ thus appeared, to alter generally accepted mechanical relations by an amount which in ordinary circumstances was imperceptible. The conditions are quite different, however, in the thermal case. Here we have not been misled by the wrong transformation formulæ, because we have already been familiar with observations at very different temperatures, and it has been impossible to make the error of neglecting the equivalent of $\left(1 - \frac{v^2}{c^2}\right)^{1/2}$, viz. $\left(1 - \frac{\tau}{\zeta}\right)^{1/4}$. That factor, therefore, automatically appeared when the relations were expressed in relativistic form.

It should be remarked also that the scope of a special theory of thermal relativity is very much more restricted than that of a special theory of mechanical relativity because the conditions justifying application of a special theory are much rarer in the former case than in the latter. The gravitational effect of one terrestrial body on the motion of another is negligible, because we do not meet with masses large enough to give it importance. On the other hand, the effect of radiation on temperature is never negligible in terrestrial matters because we never meet with a body of sufficiently great heat capacity.

There may, nevertheless, be some ways in which a special theory can do more than re-express existing knowledge in a more appropriate form. Two arbitrary constants, β and ζ , have been left in the formulæ. These determine the units of η and thermal time, respectively, and leave an arbitrariness corresponding to that introduced into mechanics by the choice of the centimetre and the second as the units of space and mechanical time. By choosing a beam of light as the "specified moving body," we can partially remove the arbitrariness of mechanical time; we then, of course, obtain unity for the velocity of light. The corresponding choice in the thermal case would be that of a body at the absolute zero for the content of S in the clock. With this choice we should similarly obtain unity for the value of the absolute zero, $-\zeta$; for dp_1 would become $-p_1$, so that we should have $\nu = -1$ and $-\zeta = 1$. The arbitrariness of β would remain, corresponding to that of the centimetre, and I think it should be possible to make a choice which would result in the

appearance of one of the constants (such as k or h) hitherto related only to the conception of atomicity in matter or energy. I have not yet, however, been able to do this.

Pending the generalization of the theory, the only other applications which it seems possible to make are to such phenomena as we know to be dependent only on temperature differences, to involve only constant temperatures, and to be independent of the properties of particular substances. These conditions, of course, limit the field very severely. The only two such phenomena that occur to me are, first, the maximum efficiency of a heat engine working between fixed temperatures; and second, the general equations of the thermoelectric circuit. What we should expect of such phenomena is that, if the relativity claimed for temperature radiation extends to them, the formulæ should be invariant under the transformation (21).

We can see at once that this is so for the first phenomenon. The simplest proof is as follows. From the last equation of (18) we have

$$\frac{\theta}{\theta_0} = \left(1 + \frac{\tau}{\xi}\right)^{1/4}, \quad \dots \quad (25)$$

where θ_0 is, of course, the ordinary absolute temperature of the arbitrarily chosen zero from which τ is measured. Any formula, therefore, which involves two absolute temperatures, θ_1 and θ_2 , only as some function of their ratio, will be invariant under the transformation (21) because θ_0 will not occur in it; we shall have

$$\frac{\theta_2}{\theta_1} = \left(\frac{1 + \tau_2/\xi}{1 + \tau_1/\xi}\right)^{1/4} \quad \dots \quad (26)$$

This is a sufficient condition for invariance, though not obviously a necessary one.

The maximum efficiency of an engine working between the constant temperatures, θ_2 and θ_1 , is

$$\epsilon = 1 - \frac{\theta_1}{\theta_2} \quad \dots \quad (27)$$

This clearly satisfies the above condition.

For the thermoelectric application consider a thermocouple consisting of two metals with junctions at constant temperatures θ_2 and θ_1 , respectively, in which the resultant E.M.F. is E . The general formulæ for such a circuit are

$$\left. \begin{aligned} e &= \theta \frac{dE}{d\theta} \\ \int (\sigma_A - \sigma_B) d\theta &= \int \theta \frac{d^2 E}{d\theta^2} d\theta \end{aligned} \right\}, \quad \dots \quad (28)$$

where e is the Peltier coefficient and σ_A , σ_B are the Thomson coefficients for the two metals. (The integral $\int (\sigma_A - \sigma_B) d\theta$ is, of course, the total Thomson effect in the circuit. Its determination does not necessarily

require the separate determination of σ_A and σ_B .) If, when we transform θ to τ in these expressions by means of the last equation of (18), θ_0 disappears, the formulæ are invariant.

We have

$$\left. \begin{aligned} \theta &= \theta_0 \left(1 + \frac{\tau}{\zeta}\right)^{1/4} \\ d\theta &= \frac{\theta_0}{4\zeta} \left(1 + \frac{\tau}{\zeta}\right)^{-3/4} d\tau \\ \frac{dE}{d\theta} &= \frac{dE}{d\tau} \frac{d\tau}{d\theta} = \frac{4\zeta}{\theta_0} \left(1 + \frac{\tau}{\zeta}\right)^{3/4} \frac{dE}{d\tau} \\ \frac{d^2E}{d\theta^2} &= \frac{4\zeta}{\theta_0^2} \left(1 + \frac{\tau}{\zeta}\right)^{1/2} \left[4\zeta \left(1 + \frac{\tau}{\zeta}\right) \frac{d^2E}{d\tau^2} + 3 \frac{dE}{d\tau} \right] \end{aligned} \right\} \dots \dots (29)$$

Hence
$$e = \theta \frac{dE}{d\theta} = 4\zeta \left(1 + \frac{\tau}{\zeta}\right) \frac{dE}{d\tau}$$

and
$$\theta \frac{d^2E}{d\theta^2} d\theta = \left[4\zeta \left(1 + \frac{\tau}{\zeta}\right) \frac{d^2E}{d\tau^2} + 3 \frac{dE}{d\tau} \right] d\tau. \dots \dots (30)$$

e and E are measured independently of temperature scales, and the formulæ, being independent of the arbitrary zero of temperature, θ_0 , are therefore invariant under the transformation (21). It may easily be verified, if desired, by direct transformation of (30) to the dashed co-ordinate system, that the expressions are actually invariant.

The great majority of thermal phenomena, however, involve change of temperature, and for this a general theory is necessary. Following the course taken by mechanical relativity, we should naturally consider the properties of "spaces" represented by a generalized fourth-power metric,

$$d\sigma^4 = g_{\alpha\beta\gamma\delta} dx^\alpha dx^\beta dx^\gamma dx^\delta, \quad [\alpha, \beta, \gamma, \delta = 1, 2] \dots \dots (31)$$

or, in our particular case,

$$d\sigma^4 = g_{1111} d\eta^4 + 4g_{1112} d\eta^3 dt + 6g_{1122} d\eta^2 dt^2 + 4g_{1222} d\eta dt^3 + g_{2222} dt^4, \quad (32)$$

and look for tensors whose vanishing would constitute the natural law of radiation of bodies and other important relations such as, for instance, the general formula for variation of specific heat, or its equivalent, with temperature. Such spaces have been studied to some extent, but as in present circumstances I am unlikely within a reasonable time to have an opportunity of mastering their properties sufficiently to make the necessary generalization, I have thought it best to publish an account of the special theory alone, in order to indicate the fundamental idea.

LIX. *Solutions of Problems Relating to Media in Contact by the Method of Wave-trains.*

By GEORGE GREEN, D.Sc., Lecturer in Applied Physics,
The University, Glasgow *.

[Received February 9, 1944.]

IN an earlier communication dealing with heat conduction problems ⁽¹⁾, a general account of the method of solution of problems by means of wave-trains is given, and illustrated by an application of the method to the particular case of concentric spherical media in contact. The summations of effects represented by wave-trains obtained in this paper are the same as would be required for any two concentric spherical media in contact and for any effects transmitted throughout the media by wave-motion. They may also be applied to effects transmitted by plane waves. The summations of wave-train effects—which are essential to facilitate the general application of the wave-train method—require to be summarized and arranged for this purpose: and it is proposed to do so in the present paper. In the later part of the present paper these tabulated results are applied to the solution of problems dealing with vibrations in composite media; this part being an extension of the work contained in a former paper ^{(2), (3)}. Thereafter attention is drawn to the fact that the method of solution by wave-trains readily provides an equation for the determination of periods of vibration of composite media under a wide variety of boundary conditions.

Consider two media in contact which extend from $x=0$ to $x=a$, and from $x=a$ to $x=b$ respectively. The fundamental problems which arise in relation to the two media are to determine the effects produced in each medium by wave-trains issuing from a source situated at any point within the system. Four cases arise according as the source is situated at either external boundary, at $x=0$, or at $x=b$; or at any point $x=x_1$ within medium I ($0 < x_1 < a$); or at any point $x=x_1$ within medium II ($a < x_1 < b$). The statement that a periodic source is located at $x=x_1$ in medium I means that two wave-trains which we may represent by $\rho_1 e^{ikt - i(\mathbf{x} - \mathbf{x}_1)\lambda}$ and $\rho_1 e^{ikt - i(\mathbf{x}_1 - \mathbf{x})\lambda}$ originate there, where the effect is $\rho_1 e^{ikt}$ on each of these trains. Each of these trains is a solution of the differential equation which governs the transmission of effects in medium I. The first train represents an effect transmitted in the positive direction of x , and the second an effect transmitted in the negative direction of x . Similarly a periodic source located at x_1 in medium II is represented by the positive and negative trains $\rho_2 e^{ikt - i\mu(\mathbf{x} - \mathbf{x}_1)\lambda}$ and $\rho_2 e^{ikt - i\mu(\mathbf{x}_1 - \mathbf{x})\lambda}$, where μ

* Communicated by the Author.

denotes $\frac{V_1}{V_2}$, the ratio of the wave-velocities, or phase velocities, of wave-trains of the same frequency $\frac{2\pi}{k}$ in the two media respectively.

Each wave-train issuing from a source undergoes reflection and transmission at an internal boundary of the system. At an external boundary it undergoes reflection. Resultant effects at any point are to be obtained by summing effects represented by each train and its continuations by reflection and transmission throughout the two media. To find an expression for the wave-systems ultimately established in medium I and medium II, arising from these original trains, we require boundary coefficients determined in general by the method indicated in the paper, reference ⁽¹⁾, p. 249, and later papers. Accordingly, following the notation employed in former papers, we take, for a train originating in medium I and initially incident upon boundary $x=a$,

A	to represent the reflection coefficient at $x=a$,
A'	„ „ transmission „ $x=a$,
B	„ „ reflection „ $x=b$,
C	„ „ retransmission „ $x=a$,
C'	„ „ reflection „ $x=a$
A ₀	„ „ reflection „ $x=0$.

In the process of summing wave-train effects due to successive reflections at boundaries $x=0$ and $x=a$, we require the summation

$$S_1 = 1/\{1 - A_0 A e^{-i2a\lambda}\}, \quad \text{or} \quad \frac{1}{S_1} = e^{-ia\lambda} E\{a\lambda, -A_0 A\}, \quad \dots \quad (1)$$

an equation defining the E function to be used below. Similarly, for the case of boundaries $x=a$ and $x=b$, we require the summation

$$S_2 = 1/\{1 - B C' e^{-i2\mu(b-a)\lambda}\}, \quad \text{or} \quad \frac{1}{S_2} = e^{-i\mu(b-a)\lambda} E\{\mu(b-a)\lambda, -B C'\}. \quad (2)$$

The summation of successive effects in medium I, arising from trains which have continued through both media and have returned to medium I, as explained in ref. ⁽¹⁾, pp. 250-1, involves the summation

$$S = \frac{1}{1 - \alpha}, \quad \text{where } \alpha = A_0 A' B C S_1 S_2 e^{-i2a\lambda}, \quad \dots \quad (3)$$

and $s = \{a + \mu(b-a)\}$.

The same summation is required in summing successive effects in medium II. Finally, in the same notation, employed below,

$$D = \frac{1}{S_1 S_2 S} \cdot \dots \quad (4)$$

Wave-train Effects.

In the statement of wave-train effects given below, two results are given for each medium termed the first effect and the total effect respec-

tively, produced in each medium by a source situated at a point within the medium. The first effect is the effect produced by the given source within the medium—regarded as isolated from the other medium. The total effect is the effect produced at any point, due to the same source, when the two media are in contact.

Case I.—Periodic source at point x_1 in medium I.

$$\text{First effect at } x \text{ in I } (x > x_1) = \frac{\rho_1}{A_0} e^{ikt} E\{x_1\lambda, A_0\} [E\{x\lambda, A_0\} S_1 - e^{ix\lambda}]. \quad (5)$$

$$\text{First effect at } x \text{ in I } (x < x_1) = \text{same expression, } x \text{ and } x_1 \text{ interchanged.} \quad (6)$$

$$\text{Total effect at } x \text{ in I } (x > x_1) = \frac{\rho_1}{A_0} e^{ikt} E\{x_1\lambda, A_0\} [E\{x\lambda, A_0\} S_1 S - e^{ix\lambda}]. \quad (7)$$

$$\text{Total effect at } x \text{ in I } (x < x_1) = \text{same expression, } x \text{ and } x_1 \text{ interchanged.} \quad (8)$$

$$\text{Total effect at } x \text{ in II} = \rho_1 e^{ikt - ix\lambda} E\{x_1\lambda, A_0\} E\{\mu(b-x)\lambda, B\} \frac{A'}{D}. \quad (9)$$

Case II.—Periodic Source at a point x_1 in medium II.

$$\begin{aligned} \text{First effect at } x \text{ in II } (x > x_1) &= \frac{\rho_2}{B} e^{ikt} E\{\mu(b-x)\lambda, B\} \\ &\times [E\{\mu(b-x_1)\lambda, B\} S_2 - e^{i\mu(b-x_1)\lambda}]. \end{aligned} \quad (10)$$

$$\text{First effect at } x \text{ in II } (x < x_1) = \text{same expression, } x \text{ and } x_1 \text{ interchanged.} \quad (11)$$

$$\begin{aligned} \text{Total effect at } x \text{ in II } (x > x_1) &= \frac{\rho_2}{B} e^{ikt} E\{\mu(b-x)\lambda, B\} \\ &\times [E\{\mu(b-x_1)\lambda, B\} S_2 S - e^{i\mu(b-x_1)\lambda}]. \end{aligned} \quad (12)$$

$$\text{Total effect at } x \text{ in II } (x < x_1) = \text{same expression, } x \text{ and } x_1 \text{ interchanged.} \quad (13)$$

$$\begin{aligned} \text{Total effect at } x \text{ in I} &= \frac{\rho_2 C}{D} e^{ikt - ix\lambda} \\ &\times E\{\mu(b-x_1)\lambda, B\} E\{x\lambda, A_0\}. \end{aligned} \quad (14)$$

Case III.—Periodic source at boundary $x=0$ in medium I.

$$\text{First effect at } x \text{ in I} = \frac{\rho_0}{A_0} e^{ikt} [E\{x\lambda, A_0\} S_1 - e^{ix\lambda}]. \quad (15)$$

$$\text{Total effect at } x \text{ in I} = \frac{\rho_0}{A_0} e^{ikt} [E\{x\lambda, A_0\} S_1 S - e^{ix\lambda}]. \quad (16)$$

$$\text{Total effect at } x \text{ in II} = \frac{\rho_0}{A_0} e^{ikt - ix\lambda} E\{\mu(b-x)\lambda, B\} \frac{A'}{D}. \quad (17)$$

Case IV.—Periodic source at boundary $x=b$, in medium II.

$$\text{First effect at } x \text{ in II} = \frac{\rho}{B} e^{ikt} [E\{\mu(b-x)\lambda, B\} S_2 - e^{i\mu(b-x)\lambda}]. \quad (18)$$

$$\text{Total effect at } x \text{ in II} = \frac{\rho}{B} e^{ikt} [E\{\mu(b-x)\lambda, B\} S_2 S - e^{i\mu(b-x)\lambda}]. \quad (19)$$

$$\text{Total effect at } x \text{ in I} = \frac{\rho}{B} e^{ikt - i\lambda x} E\{x\lambda, A_0\} \cdot \frac{C}{D}. \quad (20)$$

The products $S_2 S$ and $S_1 S$ may be written as $\frac{1}{S_1 D}$ and $\frac{1}{S_2 D}$ respectively in the above expressions. These results are necessarily symmetrical expressions with respect to the external boundaries $x=0$ and $x=b$. Thus x in medium I corresponds with $(b-x)$ in medium II: effect (5), (6), in medium I corresponds with (11), (10) in medium II respectively. Similarly (7), (8), corresponds with (13), (12), respectively, and (16) with (19).

These results are applicable directly to any two media in contact, and to either spherical or plane wave-fronts. Results included under Case I and Case II are applied to spherical media in contact, and to the subject of heat conduction in ref.⁽¹⁾. For this particular application, x and x_1 are replaced by r and r_1 respectively: A_0 takes the value -1 , and ρ_1 and ρ_2 are defined by equation (25), p. 248, of the paper, so as to represent sources of heat of unit strength per unit area, situated at the surfaces $r=r_1$ taken in medium I and medium II respectively. Similar summations may be expressed in terms of Bessel functions in the case of spherical and cylindrical media⁽⁴⁾, and these may be discussed more fully later.

Problem No. 1.

As illustrations of the application of these results we propose to consider problems involving the transmission of elastic and temperature vibrations throughout media in contact: this being an extension of the work of refs.⁽²⁾ and⁽³⁾, in which vibrations within a single uniform medium are considered. An introductory treatment of the subject of vibrations in composite media has also been given in ref.⁽⁵⁾, but a fuller treatment is desirable on account of the importance of the subject in many branches of engineering.

Taking the case of linear elastic vibrations, let u_1 and u_2 represent displacements of any layer of particles whose equilibrium position is defined by coordinate x , in a rod or column in medium I and medium II respectively, and let

$$\frac{\partial^2 u_1}{\partial t^2} = V_1^2 \frac{\partial^2 u_1}{\partial x^2}, \quad \text{and} \quad \frac{\partial^2 u_2}{\partial t^2} = V_2^2 \frac{\partial^2 u_2}{\partial x^2}, \quad (21)$$

be the equations governing the transmission of vibrations in the media respectively. In these equations, $V^2 = E/d$, where E is the Young's modulus and d the density of the medium concerned. As the particular conditions of the problem to be considered we take the case of two elastic rods in contact at section $x=a$, end $x=0$ of medium I being fixed, while end $x=b$ of medium II is free and carries a vibrator of

mass M. As initial conditions we may assume that the system starts to move from rest, the vibrator M having been displaced from its equilibrium position and then let go. We may represent these initial conditions by

$$u_1 = f_1(x), \text{ and } u_2 = f_2(x), \text{ at } t=0. \quad (22)$$

In the application of the wave-train method to determine the subsequent motion of the system, these initial conditions are to be represented ultimately by a distribution of instantaneous sources throughout medium I and medium II and at the vibrator of mass M; each instantaneous source effect being obtained from the corresponding periodic source effect by an application of Fourier's integral theorem. We accordingly require the effects produced by wave-trains:—

$\rho_1 e^{ikt - i(x-x_1)\lambda}$ and $\rho_1 e^{ikt - i(x_1-x)\lambda}$ issuing from a source at x_1 in medium I,
 $\rho_2 e^{ikt - i\mu(x-x_1)\lambda}$ and $\rho_2 e^{ikt - i\mu(x_1-x)\lambda}$ issuing from a source at x_1 in medium II,
 $\rho e^{ikt - i\mu(b-x)\lambda}$ issuing from a source at the mass M at $x=b$;

these effects being given under Case I, Case II, and Case IV respectively.

Here the value of λ is $\frac{k}{V_1}$, and the multipliers ρ_1, ρ_2, ρ are to be determined by the initial conditions at x_1 in each medium, and at the vibrator M ($x=b$), as stated above.

The boundary conditions applicable to the problem are

$$u_1 = 0 \text{ at } x=0, \text{ at all values of } t, \quad (23)$$

$$u_1 = u_2, \text{ and } E_1 \frac{\partial u_1}{\partial x} = E_2 \frac{\partial u_2}{\partial x}, \text{ at } x=a, \quad (24)$$

$$M\ddot{u}_2 = -E_2 A \frac{\partial u_2}{\partial x}, \text{ at } x=b. \quad (25)$$

In the last equation, which is the equation of motion of the vibrator M, A represents the cross-sectional area of the vibrating rod or column, assumed to be uniform and the same for both media. From these conditions, we obtain the values of the boundary coefficients applicable to the problem:—

$$A_0 = -1 : A = \frac{E_1 V_2 - E_2 V_1}{d} = -C' : \quad (26)$$

$$A' = \frac{2E_1 V_2}{d} : C = \frac{2E_2 V_1}{d} : d = E_1 V_2 + E_2 V_1. \quad (27)$$

$$B = \frac{iE_2 A + MV_2 k}{iE_2 A - MV_2 k} = -e^{i2\theta}, \text{ where } \tan \theta = \frac{E_2 A}{MV_2 k}. \quad (28)$$

$$\frac{d}{S_1} = 2e^{-ia\lambda} [E_1 V_2 \cos a\lambda + iE_2 V_1 \sin a\lambda]. \quad (29)$$

$$\frac{d}{S_2} = +2e^{i\phi} [E_2 V_1 \cos \phi - iE_1 V_2 \sin \phi], \text{ where } \phi = \{\theta - \mu(b-a)\lambda\}. \quad (30)$$

$$Dd = -4ie^{i\{\theta - a\lambda\}} [E_1 V_2 \cos a\lambda \sin \phi - E_2 V_1 \sin a\lambda \cos \phi] \\ = -4ie^{i\{\theta - a\lambda\}} \Delta. \quad (31)$$

The three periodic source effects required, whose summation values are given in Case I (7) and (9), Case II (12) and (14), Case IV (19) and (20), respectively, may now be stated in the form suited to the problem under consideration :—

$$\text{I} \left\{ \begin{array}{l} u_1(x > x_1) = i \frac{2\rho_1 e^{ikt} \sin x_1 \lambda}{\Delta} [E_1 V_2 \cos (a-x)\lambda \sin \phi \\ \quad - E_2 V_1 \sin (a-x)\lambda \cos \phi], \quad (32) \end{array} \right.$$

$$u_2 = i \frac{2\rho_1 e^{ikt} \sin x_1 \lambda}{\Delta} \sin \{\phi, x\} \cdot E_1 V_2, \quad \dots \dots \dots (33)$$

$$\text{where} \quad \{\phi, x\} = \{\theta - \mu(b-x)\lambda\}. \quad \dots \dots \dots (34)$$

$$\text{II} \left\{ \begin{array}{l} u_2(x > x_1) = i \frac{2\rho_2 e^{ikt} \sin \{\phi, x\}}{\Delta} [E_1 V_2 \cos a\lambda \sin \{\mu(x_1-a)\lambda\} \\ \quad + E_2 V_1 \sin a\lambda \cos \{\mu(x_1-a)\lambda\}]. \quad (35) \end{array} \right.$$

$$u_1 = i \frac{2\rho_2 e^{ikt} \sin x\lambda \sin \{\phi, x_1\}}{\Delta} E_2 V_1. \quad \dots \dots \dots (36)$$

$$\text{IV} \left\{ \begin{array}{l} u_2 = - \frac{\rho e^{i(kt-\theta)}}{\Delta} [E_1 V_2 \cos a\lambda \sin \{\mu(x-a)\lambda\} \\ \quad + E_2 V_1 \sin a\lambda \cos \{\mu(x-a)\lambda\}]. \quad (37) \end{array} \right.$$

$$u_1 = - \frac{\rho e^{i(kt-\theta)} \sin x\lambda}{\Delta} E_2 V_1. \quad \dots \dots \dots (38)$$

The values of ρ_1 , ρ_2 , and ρ are now to be determined so that the above expressions represent the effect—of a periodic displacement source of net amplitude $f_1(x_1)$ at x_1 in medium I,—of a periodic displacement source of net amplitude $f_2(x_1)$ in medium II,—of a periodic displacement source of net amplitude $f_2(b)$ at the vibrator M—respectively. In the cases of ρ_1 and ρ_2 , the simple relations

$$\rho_1 = f_1(x_1) : \rho_2 = f_2(x_2) \quad \dots \dots \dots (39)$$

obtain, and the sources which we have postulated at points x_1 in each medium are simple displacement sources from which wave-trains of equal amplitude issue in both directions. In the case of ρ , the source at M is one-sided, and, in addition, a special relation obtains between ρ and the effective strength of the source at M owing to the concentration of mass. If we regard M as a source of displacement of amount Xe^{ikt} per second, the corresponding rate of change of momentum of mass M is at each instant $MikXe^{ikt}$: and the equation of motion of the vibrator M regarded as a source of displacement of strength Xe^{ikt} is

$$M\ddot{u}_2 = -E_2 A \frac{\partial u_2}{\partial x} + MikXe^{ikt}. \quad \dots \dots \dots (40)$$

This equation expresses the periodic source condition—for a source of strength Xe^{ikt} : that is, this condition must be fulfilled at $x=b$ by the displacement u_2 as given in solution (37) above, where X is a known value. The condition applies, in fact, to the original wave-train issuing

from the source at the mass M , at $x=b$, as the remaining trains included in (37) satisfy boundary equation (25). This yields the relation

$$-\rho e^{-i\theta} = i \frac{MV_2 X}{R}, \quad \dots \quad (41)$$

where $R^2 = (E_2 A)^2 + (MV_2 k)^2$, and $X = 2f_2(b)$. $\dots \dots \dots$ (42)

The six expressions given in (32), (33), (35), (36), (37), (38) above, together with the evaluations of ρ_1 , ρ_2 , ρ given in (39), (41) and (42) respectively, constitute the required solution of the problem, so far as periodic sources are concerned. The two equations (37) and (38) now become, in virtue of (41) and (42),

$$u_2 = i \frac{2MV_2 f_2(b)}{R\Delta} e^{ikt} [E_1 V_2 \cos a\lambda \sin \{\mu(x-a)\lambda\} + E_2 V_1 \sin a\lambda \cos \{\mu(x-a)\lambda\}], \quad (43)$$

$$u_1 = i \frac{2MV_2 f_2(b)}{R\Delta} e^{ikt} \sin x\lambda \cdot E_2 V_1 \quad \dots \dots \dots (44)$$

respectively.

The function Δ , which appears as a denominator in all the periodic solutions, is according to its definition in (31) above a function of $a\lambda$ and $\{\theta - \mu(b-a)\lambda\}$ which might be written in the form $\Delta[a\lambda, \{\theta - \mu(b-a)\lambda\}]$. It is important to note that effects in medium II, as given in (12), (19), etc., can be expressed in terms of this function by means of the relation

$$[E\{\mu(b-x)\lambda, B\} S_2 S - e^{i\mu(b-x)\lambda}] = \frac{\Delta[a\lambda, \{-\mu(x-a)\lambda\}]}{\Delta}, \quad \dots \quad (45)$$

In the numerator θ is given the value zero, which corresponds with a boundary coefficient $B = -1$, and $(b-a)$ is replaced by $(x-a)$ as in (35) and (37) above. The relation (45) is easily proved in this case: and similar relations hold throughout the group of wave-train effects given in Cases I to IV, for First effects and Total effects.

The remaining steps required to obtain the solution of the problem proposed are:—(1) to convert the above periodic source solutions of strengths $f_1(x_1)$, $f_2(x_1)$, and $2f_2(b)$, into the corresponding instantaneous solutions respectively; (2) to integrate with respect to x_1 , from $x_1=0$ to $x_1=a$, all expressions containing $f_1(x_1)$, and to integrate with respect to x_1 , from $x=a$ to $x=b$, all expressions containing $f_2(x_1)$. The first of these steps involves integrations of the type

$$(u)_1 = \frac{1}{\pi} \int_0^\infty dk u_1, \quad \dots \dots \dots (46)$$

where $(u)_1$ represents the instantaneous source solution corresponding to the periodic source solution u_1 . This integration has to be performed for each of the six periodic solutions (32), (33), (35), (36), (43), (44). The evaluations required may be readily obtained by contour integration along a path consisting of a radial line from the origin to infinity in the sector $\left(\theta < \frac{\pi}{4}\right)$ and its image in the real axis of k , these two lines being

joined by a circular arc at infinite distance from the origin. An alternative method is to determine Cauchy's principal values for each of the integrals. In either case, for each integral, we obtain a series of terms depending on the values of k which are the roots of the equation, $\Delta=0$: and a very considerable simplification is obtained by using this equation to modify the expressions contained in (32), (35), (37), and (43). Taking, for example, the case of equation (32), we may write

$$[E_1 V_2 \cos (a-x)\lambda \sin \phi - E_2 V_1 \sin (a-x)\lambda \cos \phi] \\ = \sin x\lambda [E_1 V_2 \sin a\lambda \sin \phi + E_2 V_1 \cos a\lambda \cos \phi], \quad (47)$$

in the expression representing $u_1(x > x_1)$. By applying the same modification, with x replaced by x_1 , the expressions representing $u_1(x > x_1)$ and $u_1(x < x_1)$ become identical in form, and the integration with respect to x_1 is accordingly simplified. If we now write

$$F(k) = 2 \sin x\lambda [E_1 V_2 \sin a\lambda \sin \phi + E_2 V_1 \cos a\lambda \cos \phi] \\ \times \int_0^a f_1(x_1) \sin x_1\lambda \, dx_1, \quad (48)$$

the term of the final solution which is contributed by initial displacements represented by $f_1(x)$ is represented, for medium I, by

$$(u)_1 = - \sum_1 \frac{F(k) \cos kt}{\frac{d}{dk} \Delta}, \quad (49)$$

where the summation applies to all positive roots of equation, $\Delta=0$. In like manner, by applying the relation $\Delta=0$ to (35) or (37), we obtain

$$[E_1 V_2 \cos a\lambda \sin \mu(x_1-a)\lambda + E_2 V_1 \sin a\lambda \cos \mu(x_1-a)\lambda] \\ = \sin \{\phi, x_1\} [E_1 V_2 \cos a\lambda \cos \phi + E_2 V_1 \sin a\lambda \sin \phi], \quad (50)$$

and the expressions representing $u_2(x > x_1)$ and $u_2(x < x_1)$ become identical in form. If, in this case, we define $F(k)$ by the relation

$$F(k) = 2 \sin \{\phi, x\} [E_1 V_2 \cos a\lambda \cos \phi + E_2 V_1 \sin a\lambda \sin \phi] \\ \times \int_a^b f_2(x_1) \sin \{\phi, x_1\} \, dx_1, \quad (51)$$

we find that the term in the final solution, which depends upon the initial displacements represented by $f_2(x)$, is represented for medium II by the same expression (49), where, as before, the summation applies to all positive roots of the equation $\Delta=0$. We may employ the expression (49) above to represent the term in the final solution contributed by each of the periodic solutions (32) to (44), provided the function $F(k)$ is defined to suit each case. For (36), we take

$$F(k) = 2 \sin x\lambda \cdot E_2 V_1 \int_a^b f_2(x_1) \sin \{\phi, x_1\} \, dx_1, \quad (52)$$

and (49) then represents the term of the final solution for medium I

which depends upon the initial displacements represented by $f_2(x_1)$ for medium II. For (44) we take

$$F(k) = 2 \sin x\lambda \cdot \frac{MV^2}{R} E_2 V_1 f_2(b), \quad \dots \dots \dots (53)$$

and (49) then represents the final effect in medium I which depends upon the initial displacement of the vibrator M. The three terms determined by (48), (52), (53), inserted in (49) respectively, give by summation the total effect in medium I.

Similarly for solution (33) we take

$$F(k) = 2 \sin \{\phi, x\} E_1 V_2 \int_0^a f_1(x_1) \sin x_1 \lambda dx_1, \quad \dots \dots \dots (54)$$

and (49) then represents the term of the final solution for medium II, which depends upon the initial displacements in medium I. For (43) we take

$$F(k) = 2 \sin \{\phi, x\} [E_1 V_2 \cos a\lambda \cos \phi + E_2 V_1 \sin a\lambda \sin \phi] \times \frac{MV^2}{R} f_2(b), \quad (55)$$

and (49) represents the term of the final solution for medium II, which depends upon the initial displacement of the vibrator M. The three terms determined by (51), (54), (55), respectively, inserted in (49), give by summation the total effect in medium II.

It is worthy of note that when the distinction between media I and II is cancelled by putting $\mu=1$, $E_2=E_1=E$, and $V_2=V_1=V$, we recover the solution for the case of a uniform medium extending from $x=0$ to $x=b$, with vibrator M attached at end $x=b$: and the solution so obtained is in exact agreement with that given in refs. ^{(2), (3)}.

Problem No. 2.

Consider now, as a further example of the application of the results recorded under Cases I and II in equations (5) to (14) above, the case of two equal elastic rods in contact at section $x=a$, with the ends, $x=0$ of medium I and $x=2a$ of medium II, both fixed. As initial conditions we may assume

$$u_1 = px \quad \text{and} \quad u_2 = p(2a-x), \quad \text{at } t=0 \quad \dots \dots \dots (56)$$

equations representing an initial uniform extension in medium I and an initial uniform compression in medium II: and the problem to be solved is that of determining the subsequent displacement of any particle in medium I or medium II.

Boundary coefficients applicable to this problem are again as stated in equations (26) and (27) above, with the addition of $B=-1$ in place of (28). The periodic source solutions given under Case I (7) and (9), and under Case II (12) and (14) respectively, may now be written in the form

$$I \left\{ \begin{array}{l} u_1(x > x_1) = 2i\rho_1 e^{ikt} \sin x_1 \lambda [E_2 V_1 \sin(a-x)\lambda \cos \mu a \lambda \\ \quad \quad \quad + E_1 V_2 \cos(a-x)\lambda \sin \mu a \lambda] / \Delta, \quad (57) \\ u_2 = 2i\rho_1 e^{ikt} \sin x_1 \lambda \sin \mu(2a-x)\lambda \cdot E_1 V_2 / \Delta. \quad \dots \dots \dots (58) \end{array} \right.$$

$$\text{II} \begin{cases} u_2(x > x_1) = 2i\rho_2 e^{ikt} \sin \mu(2a-x)\lambda [E_1 V_2 \sin \mu(x_1-a)\lambda \cos a\lambda \\ \quad + E_2 V_1 \cos \mu(x_1-a)\lambda \sin a\lambda] / \Delta, \quad (59) \\ u_1 = 2i\rho_2 e^{ikt} \sin x\lambda \sin \mu(2a-x_1)\lambda \cdot E_2 V_1 / \Delta, \quad . \quad . \quad . \quad (60) \end{cases}$$

$$\text{where} \quad \Delta = E_2 V_1 \sin a\lambda \cos \mu a\lambda + E_1 V_2 \cos a\lambda \sin \mu a\lambda. \quad . \quad . \quad . \quad (61)$$

Values of S_1 and S_2 are to be obtained from (29) and (30) above, θ being given a zero value in (30).

As in previous cases, the corresponding instantaneous source solutions are to be obtained from each of these periodic source solutions by an integration of the type indicated in (46) above, each integration yielding a series of terms depending upon the values of k determined by the roots of the equation $\Delta=0$. The two processes, of integration with respect to k and integration with respect to x_1 when applied to previous solutions yield results as follows, expressed by (49) for medium I and for medium II. For medium I, if we put

$$F(k) = 2 \sin x\lambda [E_1 V_2 \sin a\lambda \sin \mu a\lambda - E_2 V_1 \cos a\lambda \cos \mu a\lambda] \\ \times \int_0^a p x_1 \sin x_1 \lambda dx_1, \quad (62)$$

then (49) represents the term in the final solution depending on the initial displacements in medium I itself: and if we put

$$F(k) = 2 \sin x\lambda E_2 V_1 \int_a^{2a} p(2a-x_1) \sin \mu(2a-x_1) dx_1, \quad . \quad . \quad (63)$$

then (49) represents the term in the final solution which depends on the initial displacements in medium II. Similarly, for medium II, if we put

$$F(k) = 2 \sin \mu(2a-x)\lambda [E_2 V_1 \sin a\lambda \sin \mu a\lambda - E_1 V_2 \cos a\lambda \cos \mu a\lambda] \\ \times \int_a^{2a} p(2a-x_1) \sin \mu(2a-x_1)\lambda dx_1, \quad (64)$$

then (49) represents the term in the final solution which depends on the initial displacements in medium II itself: and if we put

$$F(k) = 2 \sin \mu(2a-x)\lambda E_1 V_2 \int_0^a p x_1 \sin \lambda x_1 dx_1, \quad . \quad . \quad (65)$$

then (49) represents the term in the final solution which depends on the initial displacements in medium I. In all the above cases the summations refer to the positive roots of the equation $\Delta=0$, Δ being given by (61) above. When the distinction between medium I and medium II is cancelled the above results yield

$$u_1(x < a) = \sum_1^{\infty} 2 \frac{p}{a\lambda^2} \cos kt \sin x\lambda \sin a\lambda, \quad . \quad . \quad . \quad (66)$$

where the summation refers to the roots of the equation $\sin 2a\lambda=0$, a result which may readily be verified by other methods. The two solutions given above may readily be transformed so as to give the solutions of analogous problems in conduction of heat,

Problem No. 3.

As an illustration of this we may now consider a problem involving the conduction of heat in a uniform rod composed of two different materials, medium I extending from $x=0$ to $x=a$, and medium II extending from $x=a$ to $x=b$. Let the whole rod be initially at zero temperature throughout, and thereafter let the end $x=b$ be maintained at a constant temperature θ , while the end $x=0$ is maintained at zero temperature. It is required to determine the temperatures, θ_1 , at any point in medium I, and θ_2 , that at any point in medium II, at any later instant t . As in this case there is no initial distribution of heat in either medium, for the solution of the problem we require only the solutions given under Case IV, equations (18), (19), (20), above, for all of which the original wave-train issuing from the source at $x=b$ is represented by $\rho e^{ikt-i\mu(b-x)\lambda}$. The equations governing the conduction of heat in the two media are

$$\frac{d\theta_1}{dt} = \kappa_1 \frac{d^2\theta_1}{dx^2}, \quad \text{and} \quad \frac{d\theta_2}{dt} = \kappa_2 \frac{d^2\theta_2}{dx^2}, \quad \dots \quad (67)$$

respectively. From these, which correspond with (21) above, we obtain the relations between k , λ , and μ , expressed by

$$\sqrt{\frac{i k}{\kappa_1}} = i \lambda, \quad \text{or} \quad k = i \kappa_1 \lambda^2, \quad \text{and} \quad \mu = \sqrt{\frac{\kappa_1}{\kappa_2}}. \quad \dots \quad (68)$$

The boundary conditions are

$$\theta_1 = 0, \quad \text{at } x=0, \quad \text{at all values of } t, \quad \dots \quad (69)$$

$$\theta_1 = \theta_2, \quad \text{and} \quad K_1 \frac{d\theta_1}{dx} = K_2 \frac{d\theta_2}{dx}, \quad \text{at } x=a, \quad \dots \quad (70)$$

$$\theta_2 = \rho e^{ikt}, \quad \text{at } x=b. \quad \dots \quad (71)$$

The condition last stated is fulfilled by the original wave-train issuing from the source alone, so that the continuations of this train produce a zero effect at $x=b$. The condition to be fulfilled at $x=b$ is thus the same as that required at $x=0$ for all continuations of the original wave-train. From these conditions we obtain, as in previous cases, the boundary coefficients applicable to the problem

$$B=A_0=-1; \quad A = \frac{K_1 \sqrt{\kappa_2} - K_2 \sqrt{\kappa_1}}{d} = -C'; \quad \dots \quad (72)$$

$$A' = \frac{2K_1 \sqrt{\kappa_2}}{d}; \quad C = \frac{2K_2 \sqrt{\kappa_1}}{d}; \quad d = K_1 \sqrt{\kappa_2} + K_2 \sqrt{\kappa_1}. \quad \dots \quad (73)$$

It is obvious that these values are transformed into the corresponding values obtained in the preceding elastic vibration problem when K_1 , K_2 , $\sqrt{\kappa_1}$, $\sqrt{\kappa_2}$, are replaced by E_1 , E_2 , V_1 , V_2 , respectively. We may therefore obtain the values of S_1 , S_2 , and D either directly from the above coefficients or indirectly by transforming their values as stated

for the preceding problem. The periodic source solution given in Case IV, equations (19) and (20), may now be expressed in the form

$$\theta_2 = \frac{\rho e^{ikt}}{\Delta} [K_1 \sqrt{\kappa_2} \cos a\lambda \sin \mu(x-a)\lambda + K_2 \sqrt{\kappa_1} \sin a\lambda \cos \mu(x-a)\lambda], \quad (74)$$

$$\theta_1 = \frac{\rho e^{ikt}}{\Delta} K_2 \sqrt{\kappa_1} \sin x\lambda, \quad \dots \dots \dots (75)$$

where

$$\Delta = K_1 \sqrt{\kappa_2} \cos a\lambda \sin \mu(b-a)\lambda + K_2 \sqrt{\kappa_1} \sin a\lambda \cos \mu(b-a)\lambda. \quad \dots \quad (76)$$

The next step is to obtain the corresponding instantaneous source solutions by performing the integrations indicated by (46) above, with θ_2 and θ_1 respectively as integrands. In this step we may identify ρ with θ the temperature instantaneously created at end $x=b$ at $t=0$, and use λ as independent variable instead of k . If we now write

$$F_2(\lambda, t) = 2\theta\kappa_1 \lambda e^{-\kappa_1 \lambda^2 t} \Delta(x), \quad \dots \dots \dots (77)$$

where $\Delta(x)$ is the function Δ , with b replaced by \tilde{x} , and

$$F_1(\lambda, t) = 2\theta\kappa_1 K_2 \sqrt{\kappa_1} \lambda e^{-\kappa_1 \lambda^2 t} \sin x\lambda, \quad \dots \dots \dots (78)$$

we may express the solution to the problem proposed in the form

$$\theta_2 = \int_0^t dt' \left[- \sum_1 \frac{F_2\{\lambda, (t-t')\}}{\frac{d}{d\lambda} \Delta} \right], \quad \dots \dots \dots (79)$$

$$\theta_1 = \int_0^t dt' \left[- \sum_1 \frac{F_1\{\lambda, (t-t')\}}{\frac{d}{d\lambda} \Delta} \right], \quad \dots \dots \dots (80)$$

where the summations refer to all the positive roots of the equation $\Delta=0$. These integrations introduce in each case a series of constant terms whose sum can be determined by considering the steady state reached in the bar when t becomes infinitely great.

Normal Co-ordinates and Period Equation

The results tabulated above, Cases I to IV, are general summations of wave-train effects applicable to any two media in contact. These summations may be all arranged with the same denominator D , and the equation $D=0$ is therefore a general form of the equation whose roots determine the normal co-ordinates in relation to any particular problem whose solution is expressed by means of the summations of wave-train effects. From the definition of D given in equation (4) above it is clear that the roots of $D=0$ are the values of λ which make any one of the summations represented by S_1 , S_2 , S , infinite, and the geometrical progression to which it applies divergent. Its roots are therefore the roots of the three equations

$$\frac{1}{S_1} = 0, \quad \frac{1}{S_2} = 0, \quad \frac{1}{S} = 0. \quad \dots \dots \dots (81)$$

The equation is applicable to all cases of transmission of effects through a single medium, or through two media in contact. For the particular case of two elastic media in contact, together with the attached vibrator M , the periods of free vibrations of the system are determined by the roots of the equation $\Delta=0$, Δ being defined by (31) above, which may be expressed in the form

$$E_1 V_2 \tan \{\theta - \mu(b-a)\lambda\} = E_2 V_1 \tan a\lambda : \tan \theta = \frac{E_2 A}{M V_2 k}, \quad (82)$$

where θ is defined in equation (28) above. The successive roots of this equation can readily be found by a graphical process. The same equation includes a number of special cases. If the vibrator M be absent, we put $M=0$ and $\theta = \frac{\pi}{2}$, and the period equation then takes the form

$$E_1 V_2 \tan \left\{ \frac{\pi}{2} - \mu(b-a)\lambda \right\} = E_2 V \tan a\lambda. \quad (83)$$

The end point ($x=b$) in medium II is in this case "free," and the boundary condition fulfilled at ($x=b$) is $\frac{\partial u_2}{\partial x} = 0$, (25) above, which leads to the value $B=1$ for boundary coefficient B . If the end point ($x=b$) in medium II is "fixed" in its equilibrium position we may still employ (83) as the period equation. This end condition requires that $B=-1$, which is secured if $\theta=0$ in (82), and the period equation may now be written in the form

$$E_2 V_1 \tan a\lambda + E_1 V_2 \tan \mu(b-a)\lambda = 0. \quad (84)$$

In all these cases mentioned above the period equation refers to two media in contact, but in each case by a simple modification the equation can be applied to a single uniform medium extending from $x=0$ to $x=b$. Thus, by putting $E_1=E_2$, $V_1=V_2$, and $\mu=1$,

$$(82) \text{ becomes } \sin \{\theta - b\lambda\} = 0, \quad (85)$$

$$(83) \text{ becomes } \sin \left\{ \frac{\pi}{2} - b\lambda \right\} = 0, \quad (86)$$

$$(84) \text{ becomes } \sin b\lambda = 0. \quad (87)$$

Similar period equations, with alteration only in the values of fundamental constants, are obtained in corresponding problems in heat conduction and other subjects, under similar boundary conditions: the fundamental wave-trains being then the same for all.

References.

- (1) Green, G., *Phil. Mag.* vol. ix. pp. 241-260 (Feb. 1930).
- (2) Green, G. *Phil. Mag.* ser. 7, vol. xxii. pp. 1079-1088 (Dec. 1936).
- (3) Green, G., ser. 7, vol. xxix. pp. 100-102 (Jan. 1940).
- (4) Robertson, J., *Phil. Mag.* ser. 7, vol. xviii. pp. 1009-1022 (Suppl., Nov. 1934).
- (5) Lees, Dr. Charles H., *Proc. Phys. Soc.* xli. pp. 204-213 (1928-9).

LX. *Note on Certain Integrals Involving Hermite's Polynomials.*

By T. D. H. BABER and LEONID MIRSKY*.

[Received September 13, 1943.]

Introduction.

IN the present note we are concerned with the integrals

$$I_t(m, n, s) = \int_{-\infty}^{\infty} \xi^s \psi_m \frac{d^t \psi_n}{d\xi^t} d\xi,$$

where m, n, s, t are non-negative integers,

$$\begin{aligned} \psi_m &= \psi_m(\xi) = N_m e^{-\frac{1}{2}\alpha\xi^2} H_m(\xi\sqrt{\alpha}), \\ N_m &= \left(2^m \cdot m! \sqrt{\frac{\pi}{\alpha}}\right)^{-\frac{1}{2}}, \end{aligned}$$

α a given constant, and $H_m(\xi)$ the m th Hermitian polynomial. It is convenient to extend our definition of $I_t(m, n, s)$ to cases where one or more of m, n, t are negative. In that case $I_t(m, n, s)$ is defined to be zero.

Integrals of this type frequently occur in quantum-mechanical investigations. As far as we are aware, however, no method of calculating them has been published. It therefore seems worth while to indicate such a method, and also to append a table of these integrals for a few small values of s and t .

Our notation is as follows:—Greek letters denote real numbers and small Latin letters non-negative integers. aPb signifies that a and b are of the same parity.

$$J_s(m, n) = \int_{-\infty}^{\infty} H_m(\xi) H_n(\xi) e^{-\frac{1}{2}\alpha\xi^2} d\xi.$$

For $a \geq 1$,

$$n^{(a)} = \{(n+1)(n+2) \dots (n+a)\}^{\frac{1}{2}}.$$

General Formulæ.

Hermitian polynomials are defined by the development

$$e^{2\xi\eta - \eta^2} = \sum_{m \geq 0} \frac{\eta^m}{m!} H_m(\xi).$$

We therefore have

$$\sum_{m, n \geq 0} \frac{\eta^m \xi^n}{m! n!} H_m(\xi) H_n(\xi) e^{-\frac{1}{2}\alpha\xi^2} = e^{2\eta\xi} e^{-(\xi - \eta)^2} \xi^s,$$

* Communicated by the Authors.

and hence

$$\begin{aligned}
 \sum_{m, n \geq 0} \frac{\eta^m \zeta^n}{m! n!} J_s(m, n) &= e^{2\eta\zeta} \int_{-\infty}^{\infty} e^{-(\xi-\eta-\zeta)s} \xi^s d\xi, \\
 &= e^{2\eta\zeta} \int_{-\infty}^{\infty} (\theta + \eta + \zeta)^s e^{-\theta^2} d\theta \quad [\xi - \eta - \zeta = \theta], \\
 &= e^{2\eta\zeta} \sum_{0 \leq j \leq s} \binom{s}{j} (\eta + \zeta)^{s-j} \int_{-\infty}^{\infty} \theta^j e^{-\theta^2} d\theta, \\
 &= e^{2\eta\zeta} \sum_{\substack{0 \leq j \leq s \\ j \text{ even}}} \binom{s}{j} (\eta + \zeta)^{s-j} \int_0^{\infty} \phi^{\frac{j-1}{2}} e^{-\phi} d\phi \quad [\theta^2 = \phi], \\
 &= e^{2\eta\zeta} \sum_{0 \leq b \leq \frac{s}{2}} \binom{s}{2b} (\eta + \zeta)^{s-2b} \Gamma(b + \frac{1}{2}), \\
 &= \sqrt{\pi} \cdot s! \cdot e^{2\eta\zeta} \sum_{0 \leq b \leq \frac{s}{2}} \frac{2^{-2b}}{b! (s-2b)!} \sum_{0 \leq a \leq s-2b} \binom{s-2b}{a} \eta^{s-2b-a} \zeta^a, \\
 &= \sqrt{\pi} \cdot s! \sum_{\substack{a, b, c \geq 0 \\ a+2b \leq s}} \frac{2^{c-2b} \eta^{s-a-2b+c} \zeta^{a+c}}{a! b! c! (s-a-2b)!}.
 \end{aligned}$$

Equating the coefficients of $\eta^m \zeta^n$ we have

$$J_s(m, n) = \sqrt{\pi} m! n! s! \sum_{\substack{a, b, c \geq 0 \\ a+2b \leq s \\ s-a-2b+c=m \\ a+c=n}} \frac{2^{c-2b}}{a! b! c! (s-a-2b)!}.$$

Now, by reason of symmetry, it may be assumed without loss of generality that $m \geq n$, and we write $m = n + \kappa$. The conditions under the summation sign now become

$$\left. \begin{aligned} 0 &\leq a \leq n, \\ 0 &\leq a \leq \frac{s-\kappa}{2}, \\ b &= \frac{s-\kappa}{2} - a, \\ c &= n - a. \end{aligned} \right\}$$

We thus obtain *

$$\begin{aligned}
 J_s(n + \kappa, n) &= \sqrt{\pi} (n + \kappa)! n! s! 2^{n-\kappa+s} \\
 &\times \sum_{0 \leq a \leq \text{Min}(n, \frac{s-\kappa}{2})} \frac{2^a}{(n-a)! a! (a+\kappa)! \left(\frac{s-\kappa}{2} - a\right)!}.
 \end{aligned}$$

* It may be of some interest to note that

$$J_s(n + \kappa, n) = \sqrt{\pi} \cdot \frac{n! s!}{\left(\frac{s-\kappa}{2}\right)!} \cdot 2^{n-s} \times \text{coefficient of } \xi^{\frac{s+\kappa}{2}} \text{ in } (1+2\xi)^{n+\kappa} (1+\xi)^{\frac{s-\kappa}{2}}.$$

From this it follows easily that

$$I_0(n+\kappa, n, s) = 2^{\frac{\kappa}{2}-s} \alpha^{-\frac{s}{2}} s! \{n! (n+\kappa)!\}^{\frac{1}{2}} \\ \times \sum_{0 \leq a \leq \text{Min}(n, \frac{s-\kappa}{2})} \frac{2^a}{(n-a)! a! (a+\kappa)! \left(\frac{s-\kappa}{2} - a\right)!}.$$

In particular, we recognize that $I_0(m, n, s)$ will vanish unless $|m-n| \leq s$ and $m-n \leq s$. This disposes of the case $t=0$. When $t>0$ we use the well-known relation

$$H_n'(\xi) = 2n H_{n-1}(\xi),$$

and obtain

$$I_t(m, n, s) = \int_{-\infty}^{\infty} \xi^s \psi_m \frac{d^{t-1}}{d\xi^{t-1}} \frac{d\psi_n}{d\xi} d\xi \\ = \int_{-\infty}^{\infty} \xi^s \psi_m \frac{d^{t-1}}{d\xi^{t-1}} \{ \sqrt{2\alpha n} \psi_{n-1} - \alpha \xi \psi_n \} d\xi \\ = \sqrt{2\alpha n} \int_{-\infty}^{\infty} \xi^s \psi_m \frac{d^{t-1} \psi_{n-1}}{d\xi^{t-1}} d\xi \\ - \alpha \int_{-\infty}^{\infty} \xi^s \psi_m \left\{ \xi \frac{d^{t-1} \psi_n}{d\xi^{t-1}} + (t-1) \frac{d^{t-2} \psi_n}{d\xi^{t-2}} \right\} d\xi,$$

so that

$$I_t(m, n, s) = \sqrt{2\alpha n} I_{t-1}(m, n-1, s) - \alpha I_{t-1}(m, n, s+1) \\ - \alpha(t-1) I_{t-2}(m, n, s).$$

This difference equation may now be used for constructing a table of the required integrals. In doing so it should be observed that $I_t(m, n, s)$ vanishes unless both the conditions

$$|m-n| \leq s+t; \quad m-n \leq s+t$$

are satisfied. This has already been shown to be true for $t=0$ and the general result follows by induction in t , on using the above difference equation.

Table of Integrals.

We give below a list of *all* non-vanishing integrals $I_t(m, n, s)$ for which $t=0$, $0 \leq s \leq 6$, or $t=1$, $0 \leq s \leq 5$, or $t=2$, $0 \leq s \leq 4$. All results hold for $n \geq 0$.

$$I_0(n, n, 0) = 1.$$

$$I_0(n+1, n, 1) = \frac{1}{\sqrt{2\alpha}} n^{(1)}.$$

$$I_0(n, n, 2) = \frac{2n+1}{2\alpha}.$$

$$I_0(n+2, n, 2) = \frac{1}{2\alpha} n^{(2)}.$$

$$I_0(n+1, n, 3) = \frac{3(n+1)}{2\alpha\sqrt{2\alpha}} n^{(1)}.$$

$$I_0(n+3, n, 3) = \frac{1}{2\alpha\sqrt{2\alpha}} n^{(3)}.$$

$$I_0(n, n, 4) = \frac{3}{4\alpha^2} (2n^2 + 2n + 1).$$

$$I_0(n+2, n, 4) = \frac{2n+3}{2\alpha^2} n^{(2)}.$$

$$I_0(n+4, n, 4) = \frac{1}{4\alpha^2} n^{(4)}.$$

$$I_0(n+1, n, 5) = \frac{5(2n^2 + 4n + 3)}{4\alpha^2\sqrt{2\alpha}} n^{(1)}.$$

$$I_0(n+3, n, 5) = \frac{5(n+2)}{4\alpha^2\sqrt{2\alpha}} n^{(3)}.$$

$$I_0(n+5, n, 5) = \frac{1}{4\alpha^2\sqrt{2\alpha}} n^{(5)}.$$

$$I_0(n, n, 6) = \frac{5}{8\alpha^3} (4n^3 + 6n^2 + 8n + 3).$$

$$I_0(n+2, n, 6) = \frac{15}{8\alpha^3} (n^2 + 3n + 3) n^{(2)}.$$

$$I_0(n+4, n, 6) = \frac{3(2n+5)}{8\alpha^3} n^{(4)}.$$

$$I_0(n+6, n, 6) = \frac{1}{8\alpha^3} n^{(6)}.$$

$$I_1(n+1, n, 0) = -\sqrt{\frac{\alpha}{2}} n^{(1)}; \quad I_1(n, n+1, 0) = \sqrt{\frac{\alpha}{2}} n^{(1)}.$$

$$I_1(n, n, 1) = -\frac{1}{2}.$$

$$I_1(n+2, n, 1) = -\frac{1}{2} n^{(2)}; \quad I_1(n, n+2, 1) = \frac{1}{2} n^{(2)}.$$

$$I_1(n+1, n, 2) = -\frac{n+3}{2\sqrt{2\alpha}} n^{(1)}; \quad I_1(n, n+1, 2) = \frac{n-1}{2\sqrt{2\alpha}} n^{(1)}.$$

$$I_1(n+3, n, 2) = -\frac{1}{2\sqrt{2\alpha}} n^{(3)}; \quad I_1(n, n+3, 2) = \frac{1}{2\sqrt{2\alpha}} n^{(3)}.$$

$$I_1(n, n, 3) = -\frac{3(2n+1)}{4\alpha}.$$

$$I_1(n+2, n, 3) = -\frac{n+3}{2\alpha} n^{(2)}; \quad I_1(n, n+2, 3) = \frac{n}{2\alpha} n^{(2)}.$$

$$I_1(n+4, n, 3) = -\frac{1}{4\alpha} n^{(4)}; \quad I_1(n, n+4, 3) = \frac{1}{4\alpha} n^{(4)}.$$

$$I_1(n+1, n, 4) = -\frac{2n^2+16n+15}{4\alpha\sqrt{2\alpha}} n^{(1)}; \quad I_1(n, n+1, 4) = \frac{2n^2-8n-9}{4\alpha\sqrt{2\alpha}} n^{(1)}.$$

$$I_1(n+3, n, 4) = -\frac{3n+10}{4\alpha\sqrt{2\alpha}} n^{(3)}; \quad I_1(n, n+3, 4) = \frac{3n+2}{4\alpha\sqrt{2\alpha}} n^{(3)}.$$

$$I_1(n+5, n, 4) = -\frac{1}{4\alpha\sqrt{2\alpha}} n^{(5)}; \quad I_1(n, n+5, 4) = \frac{1}{4\alpha\sqrt{2\alpha}} n^{(5)}.$$

$$I_1(n, n, 5) = -\frac{15}{8\alpha^2} (2n^2+2n+1).$$

$$I_1(n+2, n, 5) = -\frac{5}{8\alpha^2} (n^2+7n+9) n^{(2)}; \quad I_1(n, n+2, 5) = \frac{5}{8\alpha^2} (n^2-n-3) n^{(2)}.$$

$$I_1(n+4, n, 5) = -\frac{4n+15}{8\alpha^2} n^{(4)}; \quad I_1(n, n+4, 5) = \frac{4n+5}{8\alpha^2} n^{(4)}.$$

$$I_1(n+6, n, 5) = -\frac{1}{8\alpha^2} n^{(6)}; \quad I_1(n, n+6, 5) = \frac{1}{8\alpha^2} n^{(6)}.$$

$$I_2(n, n, 0) = -\frac{\alpha(2n+1)}{2}.$$

$$I_2(n+2, n, 0) = \frac{\alpha}{2} n^{(2)}; \quad I_2(n, n+2, 0) = \frac{\alpha}{2} n^{(2)}.$$

$$I_2(n+1, n, 1) = -\frac{n-1}{2} \sqrt{\frac{\alpha}{2}} n^{(1)}; \quad I_2(n, n+1, 1) = -\frac{n+3}{2} \sqrt{\frac{\alpha}{2}} n^{(1)}.$$

$$I_2(n+3, n, 1) = \frac{1}{2} \sqrt{\frac{\alpha}{2}} n^{(3)}; \quad I_2(n, n+3, 1) = \frac{1}{2} \sqrt{\frac{\alpha}{2}} n^{(3)}.$$

$$I_2(n, n, 2) = -\frac{1}{4} (2n^2+2n-1).$$

$$I_2(n+2, n, 2) = n^{(2)}; \quad I_2(n, n+2, 2) = -n^{(2)}.$$

$$I_2(n+4, n, 2) = \frac{1}{4} n^{(4)}; \quad I_2(n, n+4, 2) = \frac{1}{4} n^{(4)}.$$

$$I_2(n+1, n, 3) = -\frac{1}{4\sqrt{2\alpha}} (2n^2-2n-9) n^{(1)};$$

$$I_2(n, n+1, 3) = -\frac{1}{4\sqrt{2\alpha}} (2n^2+10n+3) n^{(1)}.$$

$$I_2(n+3, n, 3) = \frac{n+8}{4\sqrt{2\alpha}} n^{(3)}; \quad I_2(n, n+3, 3) = \frac{n-4}{4\sqrt{2\alpha}} n^{(3)}.$$

$$I_2(n+5, n, 3) = \frac{1}{4\sqrt{2\alpha}} n^{(5)}; \quad I_2(n, n+5, 3) = \frac{1}{4\sqrt{2\alpha}} n^{(5)}.$$

$$I_2(n, n, 4) = -\frac{1}{8\alpha} (4n^3+6n^2-16n-9).$$

$$I_2(n+2, n, 4) = -\frac{1}{8\alpha} (n^2-13n-33) n^{(2)};$$

$$I_2(n, n+2, 4) = -\frac{1}{8\alpha} (n^2+19n+15) n^{(2)}.$$

$$I_2(n+4, n, 4) = \frac{2n+13}{8\alpha} n^{(4)}; \quad I_2(n, n+4, 4) = \frac{2n-3}{8\alpha} n^{(4)}.$$

$$I_2(n+6, n, 4) = \frac{1}{8\alpha} n^{(6)}; \quad I_2(n, n+6, 4) = \frac{1}{8\alpha} n^{(6)}.$$

Appendix.

We are indebted to Prof. P. J. Daniell for suggesting an alternative proof of our formula for $J_s(n+\kappa, n)$. It is perhaps not inappropriate to give a brief sketch of this proof.

We first use the orthogonal property of the functions $H_r(\xi)e^{-\frac{1}{2}\xi^2}$ to express ξ^s linearly in terms of Hermitian polynomials. In this manner we obtain,

$$\xi^s = s! 2^{-s} \sum_{0 \leq r \leq \frac{s}{2}} \frac{H_{s-2r}(\xi)}{r! (s-2r)!}.$$

Next we consider the integral

$$K(l, m, n) = \int_{-\infty}^{\infty} H_l(\xi) H_m(\xi) H_n(\xi) e^{-\xi^2} d\xi.$$

This is equal to the coefficient of $\lambda^l \mu^m \nu^n$ in

$$l! m! n! \int_{-\infty}^{\infty} e^{2\xi(\lambda+\mu+\nu)-\lambda^2-\mu^2-\nu^2-\xi^2} d\xi,$$

and hence we have

$$K(l, m, n) = \begin{cases} \sqrt{\pi} \cdot \frac{l! m! n! 2^{\frac{l+m+n}{2}}}{\left(\frac{-l+m+n}{2}\right)! \left(\frac{l-m+n}{2}\right)! \left(\frac{l+m-n}{2}\right)!} & \text{if } l+m+n \text{ is even and } l \leq m+n, \\ & m \leq n+l, n \leq l+m \\ 0 & \text{otherwise.} \end{cases}$$

Finally we argue as follows :

$$\begin{aligned} J_s(n+\kappa, n) &= 2^{-s} s! \sum_{0 \leq r \leq \frac{s}{2}} \frac{1}{r! (s-2r)!} \int_{-\infty}^{\infty} H_{n+\kappa}(\xi) H_n(\xi) H_{s-2r}(\xi) e^{-\xi^2} d\xi \\ &= \sqrt{\pi} 2^{n-\frac{s-\kappa}{2}} (n+\kappa)! n! s! \\ &\quad \times \sum_{\substack{0 \leq r \leq \frac{s-\kappa}{2} \\ n+r \geq \frac{s-\kappa}{2}}} \frac{2^{-r}}{r! \left(\frac{s-\kappa}{2}-r\right)! \left(n+r-\frac{s-\kappa}{2}\right)! \left(\frac{s+\kappa}{2}-r\right)!} \end{aligned}$$

(provided $\kappa \leq s$, $s-\kappa$ even)

$$= \sqrt{\pi} 2^{n-s+\kappa} (n+\kappa)! n! s! \sum_{0 \leq a \leq \min(n, \frac{s-\kappa}{2})} \frac{2^a}{(n-a)! a! (a+\kappa)! \left(\frac{s-\kappa}{2}-a\right)!}$$

on putting

$$a = \frac{s-\kappa}{2} - r.$$

LXI. *Solution of the Restricted Problem of the Random Walk.*

By LUDWIK SILBERSTEIN, Ph.D.*

[Received August 9, 1943.]

THE problem of the Random Walk, proposed under this name by Karl Pearson in 1905 ('Nature,' vol. lxxii. p. 294), but investigated already in 1880 and 1889 by Lord Rayleigh † in connection with the composition of iso-periodic vibrations of equal amplitude and of phases distributed at random, may be restated as follows :—

A particle, moving in a plane, starts from a point O and makes n rectilinear steps of equal length, but of haphazardly varying direction. All directions being equally likely, it is required to find the probability that, at the end of this journey, the particle shall be at a distance r to $r+dr$ from O.

Rayleigh first solved this problem (*loc. cit.*) only approximately, for a great number n of steps. Later on, in 1919, he gave a formal solution for any n (even in three, instead of two dimensions) in the shape of an integral ‡, which he evaluated completely, by an ingenious method, for $n=3$ up to 6, inclusively. The probability is a rational function of r , of increasing complexity as to the structure of its numerical coefficients. It consists, moreover, of several branches of different form (as might be expected), valid for different ranges of r . Thus, *e. g.*, for six steps, say unit steps, the probability function has one form, the simplest, for $0 < r < 2$, another, somewhat more complicated, for $2 < r < 4$, and a third for $4 < r < 6$. By Rayleigh's method it is theoretically possible to evaluate his integral for any desired number of steps. But the computation becomes, with increasing n , extremely tedious, and since, even apart from this practical difficulty, the probability function changes its form with varying n and splits into an ever growing number of branches, the construction of a unique formula valid for any desired n and any distance $r \leq n$, is manifestly out of the question.

Such being the case, it has occurred to me to consider what is called in the title the problem of the "restricted" random walk.

Namely, let the particle be constrained to take its (unit) steps only in two fixed directions §, say $\pm \mathbf{i}$ and $\pm \mathbf{j}$, so that the position vector of the particle after n steps has any of the values

$$\mathbf{r} = (a - a')\mathbf{i} + (b - b')\mathbf{j} = x\mathbf{i} + y\mathbf{j},$$

* Communicated by the Author.

† 'Scientific Papers,' vol. i. p. 491 ; iv. p. 370, and a short note in vol. v. p. 256. Also 'Theory of Sound,' vol. i. pp. 36-42 (2nd ed., 1894).

‡ Phil. Mag. vol. xxxvii. p. 321 (1919) : 'Scientific Papers,' vol. vi. p. 604.

§ Whether orthogonal or oblique is manifestly immaterial. Our wording and figures will be such, as if these directions were orthogonal.

where a , etc., are any positive integers (including zero) satisfying the condition

$$a + a' + b + b' = n.$$

All possible end-points (x, y) of the particle's journey form a discrete array or lattice of points spread over a square, namely, $n+1$ rows of $n+1$ points each, and thus a total of

$$N = (n+1)^2 \quad \dots \dots \dots (1)$$

Fig. 1.

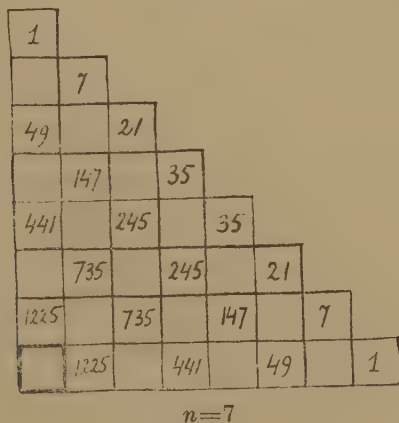
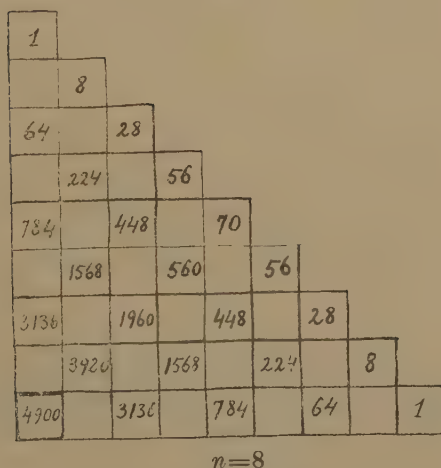


Fig. 2.



points. For $n=7$ and $n=8$, these points are represented in fig. 1 and fig. 2 by numbered squares. (The meaning of the numbers will be explained presently.) Since their distribution is symmetrical with respect to the axes i, j , a quadrant only is shown in either figure.

The position of these points themselves is readily ascertained, whatever the value of n . The problem consists in determining the probability,

$p_n(x, y)$, of reaching any one of the points (x, y) in n steps. We shall see that this probability can be expressed by a single arithmetical function, valid for any number of steps and for any lattice point (x, y) .

Let $f_n(x, y)$ be the number of ways of reaching a point (x, y) in n steps. The total number of equally probable "ways" being 4^n , the required probability is

$$p_n(x, y) = \frac{1}{4^n} f_n(x, y). \quad (2)$$

The problem is thus reduced to finding all numbers $f_n(x, y)$ which we shall call briefly *the frequencies* of the lattice points. For the first few values of n these frequencies are readily computed from the frequencies for the next lower n . The numbers attached to the "points" (squares) in fig. 1 and fig. 2 give just their frequencies computed, originally, in this manner for the sake of a "first orientation." It will be seen from these figures that even for moderate values of n the distribution of frequencies among the lattice points is rather complicated.

Our task is to find a general expression for all these frequencies, valid rigorously for any n .

The only way of accomplishing it is, actually, by adding up the frequencies of the four neighbours of (x, y) in the lattice for the next lower n , that is to say, by solving the functional equation

$$f_{n+1}(x, y) = f_n(x+1, y) + f_n(x-1, y) + f_n(x, y+1) + f_n(x, y-1). \quad (3)$$

and by determining the arbitrary constant factor of the solution from a single known frequency, such as $f_1(1, 0) = 1$ or $f_2(1, 1) = 2$.

The equation holds for all lattice points. For any other point the frequency f vanishes, by definition. The complete lattice, consisting of $(n+1)^2$ points, is limited by the four border rows of points, $x+y = \pm n$ and $x-y = \pm n$. It can be exhausted by the $n+1$ rows $x+y = n-2\kappa$ ($\kappa = 0, 1, 2, \dots, n$) or, as well, by the $n+1$ rows $x-y = n-2\kappa$, each row containing $n+1$ points. Manifestly, we have a lattice point when and only when the three numbers $n, x+y, x-y$ are, simultaneously, all even or all odd.

Instead of x, y let us introduce the new co-ordinates

$$\sigma = x+y, \quad \delta = x-y,$$

so that the boundary of the lattice will consist of the four rows $\sigma = \pm n$, $\delta = \pm n$, and the lattice points will be characterized by n, σ, δ all even or all odd. In this respect, then, the new co-ordinates are more simple than x, y and can be expected to offer other advantages.

In terms of σ, δ , the functional equation for the frequency becomes

$$f_{n+1}(\sigma, \delta) = f_n(\sigma+1, \delta+1) + f_n(\sigma-1, \delta-1) + f_n(\sigma+1, \delta-1) + f_n(\sigma-1, \delta+1).$$

Let us try to satisfy it by the product of a function of n and σ alone into a function of n and δ only, say

$$f_n(x, y) = g_n(\sigma) h_n(\delta).$$

Then

$$\begin{aligned} g_{n+1}(\sigma)h_{n+1}(\delta) &= g_n(\sigma+1)[h_n(\delta+1)+h_n(\delta-1)] \\ &\quad + g_n(\sigma-1)[h_n(\delta-1)+h_n(\delta+1)] \\ &= [g_n(\sigma+1)+g_n(\sigma-1)][h_n(\delta+1)+h_n(\delta-1)], \end{aligned}$$

which splits into

$$g_{n+1}(\sigma)=g_n(\sigma+1)+g_n(\sigma-1) \quad \text{and} \quad h_{n+1}(\delta)=h_n(\delta+1)+h_n(\delta-1).$$

In fine, the co-ordinates σ, δ have the remarkable property of "separating the variables" (which the original co-ordinates x, y do not possess).

We now have, for both g_n and h_n , the simple equation

$$\phi_{n+1}(s)=\phi_n(s+1)+\phi_n(s-1),$$

which is, for instance, the functional equation for a Random Walk in one dimension (unit steps back and forth on a line), $\phi_n(s)$ being the number of ways of reaching a distance s from the origin in n steps. The well-known solution of this simple problem is

$$\phi_n(s)=\binom{n}{\frac{n+s}{2}}=\frac{n!}{\left(\frac{n+s}{2}\right)!\left(\frac{n-s}{2}\right)!}.$$

We have, therefore,

$$f_n(x, y)=c \binom{n}{\frac{n+\sigma}{2}} \binom{n}{\frac{n+\delta}{2}}.$$

The, so far arbitrary, constant factor c may be determined from any particular frequency, say from $f_2(1, 1)=2$, which gives

$$2=c \binom{2}{0} \binom{2}{1}=2c, \quad c=1.$$

Thus the required frequency of any lattice point is

$$f_n(x, y)=\binom{n}{\frac{n+x+y}{2}} \binom{n}{\frac{n+x-y}{2}} \quad \dots \quad (4)$$

or, in terms of factorials,

$$f_n(x, y)=\frac{(n!)^2}{\left(\frac{n+x+y}{2}\right)!\left(\frac{n-x-y}{2}\right)!\left(\frac{n+x-y}{2}\right)!\left(\frac{n-x+y}{2}\right)!}.$$

This is the complete solution of the problem.

So far, the frequency. The probability of reaching a point (x, y) in n steps is, by (2) and (4),

$$p_n(x, y)=\frac{1}{4^n} \binom{n}{\frac{n+\sigma}{2}} \binom{n}{\frac{n+\delta}{2}}, \quad \dots \quad (5)$$

where $\sigma=x+y$ and $\delta=x-y$ or $y-x$, indifferently. This probability is

an even function of x and of y . Also, $p_n(y, x) = p_n(x, y)$, as could be expected. The frequencies of all axial points are exact squares, *e. g.*,

$$f_n(x, 0) = \left[\frac{n}{2} \right]^2,$$

and the frequency of border points, such as $x + y = n$, has the simple expression

$$f_n(x, n-x) = \binom{n}{x}.$$

Since, for every lattice point, n , σ , δ are all even or all odd, the numbers $\frac{1}{2}(n+\sigma)$, $\frac{1}{2}(n+\delta)$, appearing in (4), (5), are always integers. Needless to say, the sum of all $(n+1)^2$ frequencies as expressed by (4) must be equal to 4^n , which gives one more identity concerning aggregates of the binomial coefficients.

The formulæ (4) and (5) hold rigorously for any number of steps and for all lattice points. It may be interesting to see the form to which the probability function tends with ever increasing n .

In view of the noted symmetries it is enough to consider positive values of x , y . Then, if n be a great number, so is also $n+x+y$, but not necessarily $n-x-y$ and the remaining two numbers appearing through the factorials, unless we assume that

$$\frac{x}{n} \text{ and } \frac{y}{n} \text{ are small fractions,}$$

or, at least, not approaching unity very closely. If this be granted, then, together with n itself, $n \pm \sigma$ and $n \pm \delta$ will be great numbers, and their factorials may be replaced by Stirling's asymptotic expression. Thus the probability (5) becomes

$$p_n(x, y) = \frac{2}{\pi n} \cdot \left[\left(1 - \frac{\sigma^2}{n^2} \right) \left(1 - \frac{\delta^2}{n^2} \right) \right]^{-\frac{1}{2}} \\ \times \left[\left(1 - \frac{\sigma}{n} \right)^{\frac{n-\sigma}{2}} \left(1 + \frac{\sigma}{n} \right)^{\frac{n+\sigma}{2}} \left(1 - \frac{\delta}{n} \right)^{\frac{n-\delta}{2}} \left(1 + \frac{\delta}{n} \right)^{\frac{n+\delta}{2}} \right]^{-1}.$$

whence, up to terms of the third order in σ/n or δ/n ,

$$\log \left(\frac{\pi n}{2} p \right) = \frac{\sigma^2 + \delta^2}{2n^2} - \frac{\sigma^2 + \delta^2}{2n} + \frac{\sigma^2 + \delta^2}{2n} - \frac{x^2 + y^2}{n}.$$

Thus, the asymptotic value of the probability, valid under the stated conditions, is

$$p_n(x, y) = \frac{2}{\pi n} e^{-\frac{x^2 + y^2}{n}} \dots \dots \dots (5a)$$

Notice that the rigorous functional equation for the probability which, by (3) and (2), is

$$p_{n+1}(x, y) = \frac{1}{4} \{ p_n(x+1, y) + p_n(x-1, y) + p_n(x, y+1) + p_n(x, y-1) \},$$

may be replaced under the present circumstances by the differential equation

$$\frac{\partial p}{\partial n} = \frac{1}{4} \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} \right),$$

and that this equation is, in fact, satisfied by (5 a).

If the axes i, j happen to be orthogonal, then $x^2 + y^2 = r^2$ is the squared distance of the lattice point (x, y) from the origin. The result (5 a), however, holds for any oblique axes as well, when $x^2 + y^2$ has not this geometrical meaning. Keeping this in mind, we may note that (5 a) coincides, apart from a numerical factor, with Rayleigh's approximate solution of the *unrestricted* two-dimensional problem ('Theory of Sound,' i. p. 40), found by integrating the last quoted equation.

The rigorous probability (5) can be written

$$\frac{1}{2n} \left[\frac{n}{\frac{n+\sigma}{2}} \right] \cdot \frac{1}{2n} \left[\frac{n}{\frac{n+\delta}{2}} \right]$$

$$\text{or} \quad p_n(x, y) = P_n(\sigma) \cdot P_n(\delta) = P_n(x+y) \cdot P_n(x-y). \quad (5')$$

It is thus equal to the product of the probabilities of reaching, always in n steps, the points σ and δ in two independent one-dimensional random walks. It is precisely this separability of variables, in the "canonical" co-ordinates $\sigma = x+y$, $\delta = x-y$ *, which has enabled us to find a simple solution of the "two-dimensional" problem, that is to say, for a point lattice or network consisting of quadrangular meshes. (In fact, what is relevant is not "dimensionality," but the structure of the implied network.) No such separation of variables in the functional equation seems to be possible for other networks, including the simplest cubic lattice (x, y, z) , as the reader may find by trying it.

129, Seneca Parkway,
Rochester, N.Y.
July 10, 1943.

* The original variables x, y are not, rigorously, canonical co-ordinates. They only become so, approximately, for great n , etc. For, since the asymptotic value of the product $P(\sigma)P(\delta)$ happens to contain σ and δ only through the exponential of $\sigma^2 + \delta^2 = 2(x^2 + y^2)$, the probability becomes, as in (5 a), $p_n(x, y) = P_n(x) \cdot P_n(y)$.

LXII. *The Measurement of X-ray Wave-lengths by the Powder Method :
CrK β_1 and MnK β_1 .*

By H. LIPSON and LILLIAN E. R. ROGERS, Cavendish Laboratory,
Cambridge *.

[Received June 26, 1944.]

Introduction.

THE wave-lengths of the principal X-ray spectrum lines of the elements from titanium (22) to germanium (32) have been accurately measured by many different investigators, and a summary of their results is given in a table prepared by Bearden and Shaw ⁽¹⁾. This table shows a general agreement of the order of 0.001 per cent. for characteristic radiations in common use, but there is one marked exception: the wave-length of the MnK β_1 is given as 1906.195 X by Eriksson ⁽²⁾ and 1906.301 X and 1906.274 X by Bearden and Shaw ⁽¹⁾, the difference being about 0.005 per cent.

We have made an attempt to measure this wave-length by the powder method. Recent work has shown that an accuracy approaching 0.001 per cent. can be obtained from powder photographs in the comparison of linear dimensions, the determinations of the axial ratio of quartz ⁽³⁾ being consistent to this accuracy. In order to test the method we have also used it to re-measure the wave-length of CrK β_1 , the determinations of this by previous workers ^(1, 2) being in quite good agreement.

The powder method has the great advantage over the methods used by the earlier workers in that no elaborate apparatus need be used. This is due to two reasons. First, in the measurement of orders of diffraction with Bragg angles near 90°, errors in determination of θ are relatively unimportant; for instance, an error of 1 minute of arc at 84° is equivalent to an error of two seconds at 20°. Thus an accurately divided scale is not required. Secondly, systematic errors, such as those due to mis-setting of the specimen, can be eliminated and so no extreme precautions are necessary on this account.

Experimental Method.

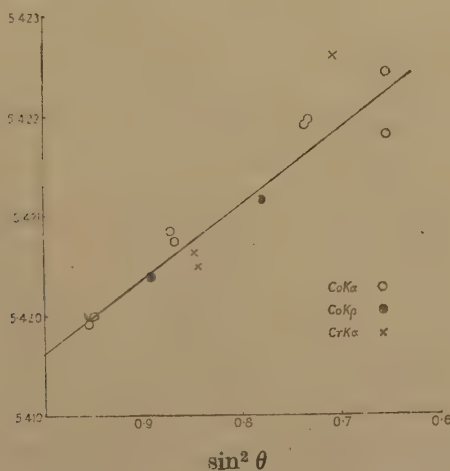
The method we have devised involves the taking of a powder photograph of a cubic substance with a mixture from the same target of the two radiations to be compared; the lattice parameter can then be evaluated in terms of the known wave-lengths and the unknown can be determined from this.

* Communicated by the Authors.

For $\text{CrK}\beta_1$ silicon was chosen as a specimen because it gives a reflection (333) at 86° . It was decided to use an element for the purpose because the reflections given by alloys are not usually quite as sharp. The silicon was ground to a fine powder and annealed to remove strain; a thin cylindrical rod was then made by mixing the powder with powdered gum tragacanth, wetting the mixture with saliva, and rolling out between the fingers the paste formed after a few minutes.

Powder photographs were taken in a camera of 19 cm. diameter, adapted to take one film which received only the reflections with θ greater than about 50° ; by mounting the film in this way no calibration of the camera is required (4).

The radiation was a mixture of CrK with either CuK or CoK . For the former mixture a copper target was plated with chromium and was then rubbed with emery until it gave a reasonable quantity of both radiations when in the X-ray tube. For the latter a similar process



Extrapolation curve for the lattice parameter of silicon with Co and Cr radiations.

was used except that the chromium was plated on to a plated layer of cobalt. $\text{CuK}\alpha$ and $\text{CoK}\alpha$ radiations were chosen as standards because they also give reflections from silicon at high angles; their wave-lengths were taken as the means of the values in the table given by Bearden and Shaw—1537.40 X for $\text{CuK}\alpha_1$, 1541.23 X for $\text{CuK}\alpha_2$, 1785.29 X for $\text{CoK}\alpha_1$ and 1789.17 X for $\text{CoK}\alpha_2$.

The lattice parameter of the specimen was found by the usual extrapolation method (5), each value being corrected for refraction. A typical plot is shown in the figure. In this figure the radii of the circles representing the results from CoK radiation have been made equivalent to 0.00005 ÅX, and it will be seen that a reasonable straight line through the points intersects these circles above $\sin^2 \theta = 0.9$. It can therefore be claimed

that an accuracy approaching 0.00005 kX, or 0.001 per cent., has been achieved.

The value of $\sin^2\theta$ for line 333 with $\text{CrK}\beta_1$ radiation is 0.995. The apparent lattice parameter, a , at this value was read from the extrapolation curve, and was inserted in the equation

$$\lambda = 2a \sin \theta / \sqrt{27}.$$

By these methods all systematic errors are eliminated except those due to differences in absorption. The absorption coefficient should be greater for the longer wave-lengths, and it can be shown that this should make the extrapolation line lower for these wave-lengths. The effect, however, is probably small, and since the points for the other wave-lengths lie reasonably well on the same line there is no reason to suppose that the points for $\text{CrK}\beta_1$, whose wave-length lies between the others, should not agree also.

Similar methods were applied for $\text{MnK}\beta_1$. The same specimen could be used, as the reflection 440 has a Bragg angle of 84° . A photometer

TABLE I.

Comparison radiation.	Lattice parameter. kX.	Wave-length. X.
Cu	5.41959	2080.57
Cu	5.41960	2080.61
Co	5.41961	2080.63
Co	5.41966	2080.65
Co	5.41963	2080.57
Co	5.41963	2080.61
Mean....	5.41962	2080.61

curve of this reflection on one of the photographs showed that its half-peak width was 1.02 X. Within the experimental error this is equal to the value 1.07 X obtained by Roseberry and Bearden⁽⁶⁾ and claimed by them to represent the natural width of the line; it therefore appears that any instrumental broadening of the powder line must be small. Bradley and Jay⁽⁵⁾ have suggested that the curvature of a high-order line should result in a displacement because it is composed of several overlapping lines, but the sharpness of the line we have measured shows that there is no appreciable effect in this case

The Wave-length of $\text{CrK}\beta_1$.

Six photographs were taken and the results are shown in Table I. The value of the lattice parameter derived from each photograph is also given in this table. Although no attempt was made to keep the temperature constant, it was observed that for all the photographs the temperature of the laboratory was between 18 and 19°C ; the variations

in the lattice parameter are therefore probably of no physical significance. The mean value is in good agreement with the values determined by Eriksson⁽²⁾ (2080.586 X) and Bearden and Shaw⁽¹⁾ (2080.597 X and 2080.587 X). The value of the wave-length may therefore be taken as 2080.60 ± 0.01 X.

The Wave-length of $\text{MnK}\beta_1$.

Five photographs were taken and the results are shown in Table II., the films being measured by each of us independently.

At first sight it appears that the results are inconclusive; the mean value lies between the two previously determined values and the spread, 0.22 X, is much greater than was expected. In view, however, of the fact that this spread is so much greater than the amount, 0.08 X, obtained in the measurement of $\text{CrK}\beta_1$, we felt that a closer examination of the results was required to see if the variation had any real significance.

TABLE II.

Comparison radiation.	Lattice parameter. kX.	Wave-length. X
Cu	5.41959	1906.28
	5.41981	1906.34
Cu	5.41973	1906.39
	5.41952	1906.34
Co	5.41960	1906.26
	5.41961	1906.25
Co	5.41965	1906.17
	5.41946	1906.18
Co	5.41963	1906.18
	5.41967	1906.24
Mean	5.41962	1906.26

It will be seen from Table II. that the results can be divided into two sets: those obtained by comparison with $\text{CuK}\alpha$ and those obtained by comparison with $\text{CoK}\alpha$. The former set is included in the range 1906.34 ± 0.06 X and the latter in the range 1906.21 ± 0.05 X. These two ranges do not overlap and the variations are now of the same order as those given by the results for $\text{CrK}\beta_1$. It therefore seems that some variation is possible in the wave-length of $\text{MnK}\beta_1$, and we suggest that this is the reason for the discrepancy between the results of the previous workers. Furthermore, the mean value of our first set of results is in good agreement with Bearden and Shaw's value, and that of the second set with Eriksson's. It is possible, therefore, that only two wave-lengths occur, but much more work would be necessary in order to establish this. The two wave-lengths would have values 1906.31 ± 0.02 X and 1906.20 ± 0.01 X.

Discussion of the Results for $\text{MnK}\beta_1$.

The inconstancy of the wave-length of $\text{MnK}\beta_1$ was unexpected, but a consideration of the production of the K spectra of the transition elements of the first long period renders the result less surprising. The $\text{K}\beta_1$ line is produced when an electron falls from the M shell of an atom to the K shell, and the M shell of these transition elements is the unfilled one. If there is any change in the content of the M shell it is probable that the mean energy of the electrons in that shell would change; thus the wave-length, which is inversely proportional to the difference of energy of the M and K shells, would also change.

What is surprising is that an electronic shell of a pure element should be capable of variation. Manganese, however, is an unusual element in that it is capable of existing in different crystalline forms at room-temperature. The stable form contains four sets of atoms of different effective atomic radii, and Hume-Rothery⁽⁷⁾ has suggested that each set consists of atoms with different numbers of electrons in the M shells. If this is so, the metastable forms will presumably be different in this respect. It is known that manganese can be electrodeposited with different structures⁽⁸⁾ and we hope to undertake further experiments to see if the change in wave-length is associated with change in structure.

Although the use of $\text{MnK}\beta_1$ for the accurate determination of lattice parameters is obviously ruled out, some decision is necessary as to what value should be inserted in standard tables of wave-lengths. Our experience suggests that the larger value (1906.31 X) is the more likely to occur in practice, and this is also in good agreement with the value 1906.29 X given by interpolation between the $\text{K}\beta_1$ wave-lengths of the neighbouring transition elements.

The Lattice Parameter of Silicon.

The work has led incidentally to a reliable determination of the lattice parameter of silicon, and it is probably worth while to record it here. The material was of 99.85 per cent. purity, but since the chief impurity was oxygen, and was presumably present at SiO_2 , it is probable that the lattice parameter is representative of that of silicon of more than 99.9 per cent. purity.

The results for the eleven photographs recorded in Tables I. and II. give a mean value of the lattice parameter of 5.41962 kX with a mean deviation of 0.00005 kX. All the photographs were taken at temperatures between 18 and 19° C. This result is in excellent agreement with that of Jette and Foote⁽¹⁰⁾, who give 5.41982 ± 0.00003 kX at 25° C.; if we take the thermal expansion as $7 \times 10^{-6(9)}$ our result at 25° C. would be 5.41987 kX. We may therefore take the lattice parameter at 5.41984 ± 0.00005 kX. at 25° C., the limits of error being determined as much by the error in the X-ray wave-lengths as by the internal consistency in the measurements,

Summary.

An attempt has been made to measure the wave-lengths of $\text{CrK}\beta_1$ and $\text{MnK}\beta_1$ by means of the powder method. An accurate value of the lattice parameter of a cubic material (silicon) was found in terms of the wave-lengths of $\text{CuK}\alpha$ and $\text{CoK}\alpha$ radiations and the two wave-lengths were then found in terms of this lattice parameter.

For $\text{CrK}\beta_1$ the value $2080.61 \pm 0.04 \text{ kX}$ was obtained and by combination with previous results it is suggested that the best value is $2080.60 \pm 0.01 \text{ X}$. The wave-length of $\text{MnK}\beta_1$ appeared to be capable of some variation. Differing values have been obtained depending on whether the manganese is plated on copper or on cobalt. The value in the former case, $1906.31 \pm 0.02 \text{ kX}$ seems to be the one most likely to occur in practice. A possible reason for the variation in wave-length is suggested.

The value of the lattice parameter of silicon is derived as $5.41964 \pm 0.00005 \text{ kX}$ at 25°C .

Acknowledgement.

We wish to thank Professor Sir Lawrence Bragg for his interest in this work.

References.

- (1) Bearden and Shaw, *Phys. Rev.* *xlvi*. p. 18 (1935).
- (2) Eriksson, *Zeits. f. Physik*, *xlvi*. p. 360 (1928).
- (3) Wilson and Lipson, *Proc. Phys. Soc.* *liii*. p. 245 (1941).
- (4) Lipson and Wilson, *J. Sci. Instr.* *xviii*. p. 144 (1941).
- (5) Bradley and Jay, *Proc. Phys. Soc.* *xlv*. p. 563 (1932).
- (6) Roseberry and Bearden, *Phys. Rev.* *i*. p. 207 (1936).
- (7) Hume-Rothery, *Phil. Mag.* (7) *xi*. p. 662 (1931).
- (8) Bradley, *Phil. Mag.* (6) *i*. p. 1024 (1925).
- (9) van Arkel, 'Reine Metalle,' p. 552. (Berlin, Springer, 1939.)
- (10) Jette and Foote, *J. Chem. Phys.* *iii*. p. 605 (1935).

LXIII. *Two-Dimensional Compressible Flow past a Solid Body in Unlimited Fluid or Symmetrically Placed in a Channel.*

By S. GOLDSTEIN, F.R.S., and M. J. LIGHTHILL, B.A., of the
Aerodynamics Department, N.P.L.*

[Received February 17, 1944.]

Summary.

The Janzen⁽¹⁾-Rayleigh⁽²⁾ process for finding the first-order effect of compressibility on the potential flow past a solid body consists in expanding the velocity potential, stream function or velocity components

* Communicated by E. F. Relf, F.R.S.

in powers of M^2 , and determining the terms in M^2 , where M is the Mach number. The problem is solved explicitly in this report for two-dimensional flow past a symmetrical body at zero incidence, either in unlimited fluid or symmetrically placed in a channel, when the irrotational incompressible flow is known. The asymmetric case is also briefly considered.

According to the linear perturbation theory⁽³⁾ the velocity increment at, for example, the position of maximum thickness of the body over the velocity of the undisturbed stream should be multiplied by $(1-M^2)^{-\frac{1}{2}}$ in unlimited fluid to allow for the effects of compressibility; the additional increment due to the constraint of the channel-walls is to be multiplied by $(1-M^2)^{-3/2}$ for a wide channel. To order M^2 these factors become $1+(\frac{1}{2})M^2$, $1+(\frac{3}{2})M^2$ respectively. The results of the linear perturbation theory should be correct to the first order in the thickness of the body, and it is shown that the solution here obtained agrees with these results to the first order in M^2 and in the thickness of the body.

1. Introduction, Equations for the Stream Function.

CONSIDER the steady two-dimensional irrotational isentropic flow of a compressible fluid past a solid body, the fluid being either infinite in extent or bounded by the walls of a channel. Take an origin of rectangular Cartesian co-ordinates inside the body; if the body is in a channel take the axis of x parallel to the walls of the channel. Let the undisturbed velocity of the fluid be U_0 parallel to the axis of x ; denote by U , V the components of the fluid velocity at any point, and by p , ρ and c the pressure, density and velocity of sound, respectively, their values in the undisturbed stream being p_0 , ρ_0 and c_0 . Let γ be ratio of the specific heats, and Q the resultant fluid velocity. Then

$$P + \frac{1}{2}Q^2 = \text{constant}, \quad \dots \dots \dots (1)$$

where
$$P = \int \frac{dp}{\rho} = \frac{\gamma}{\gamma-1} \frac{p}{\rho} = \frac{c^2}{\gamma-1} = \frac{c_0^2}{\gamma-1} \left(\frac{\rho}{\rho_0} \right)^{\gamma-1} \dots \dots \dots (2)$$

Hence
$$c^2 = c_0^2 + \frac{1}{2}(\gamma-1)(U_0^2 - Q^2) \dots \dots \dots (3)$$

and
$$\frac{\rho_0}{\rho} = \{1 + \frac{1}{2}(\gamma-1)M^2(1 - Q^2/U_0^2)\}^{-1/(\gamma-1)}, \quad \dots \dots \dots (4)$$

where
$$M = U_0/c_0. \quad \dots \dots \dots (5)$$

The equation of continuity is satisfied by putting

$$\rho U = \rho_0 U_0 \frac{\partial \psi}{\partial y}, \quad \rho V = -\rho_0 U_0 \frac{\partial \psi}{\partial x}, \quad \dots \dots \dots (6)$$

and from the equation

$$\frac{\partial U}{\partial y} = \frac{\partial V}{\partial x} \quad \dots \dots \dots (7)$$

for the absence of vorticity we then obtain

$$\rho_0 U_0 \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} \right) = U \frac{\partial \rho}{\partial y} - V \frac{\partial \rho}{\partial x} = \frac{\rho U}{c^2} \frac{\partial P}{\partial y} - \frac{\rho V}{c^2} \frac{\partial P}{\partial x}, \quad (8)$$

since
$$dP = \frac{dp}{\rho} = c^2 \frac{d\rho}{\rho} \quad (9)$$

Hence
$$\begin{aligned} \nabla^2 \psi &= \frac{1}{c^2} \left(\frac{\partial \psi}{\partial x} \frac{\partial P}{\partial x} + \frac{\partial \psi}{\partial y} \frac{\partial P}{\partial y} \right) \\ &= -\frac{1}{2c^2} \left(\frac{\partial \psi}{\partial x} \frac{\partial Q^2}{\partial x} + \frac{\partial \psi}{\partial y} \frac{\partial Q^2}{\partial y} \right) \end{aligned} \quad (10)$$

with c^2 given by (3).

If now we suppose ψ expanded in powers of M^2 ,

$$\psi = \psi_0 + M^2 \psi_1 + \dots, \quad (11)$$

and write
$$K = q^2 = \left(\frac{\partial \psi_0}{\partial x} \right)^2 + \left(\frac{\partial \psi_0}{\partial y} \right)^2, \quad (12)$$

then
$$\nabla^2 \psi_0 = 0, \quad (13)$$

$$\nabla^2 \psi_1 = -\frac{1}{2} \left(\frac{\partial \psi_0}{\partial x} \frac{\partial K}{\partial x} + \frac{\partial \psi_0}{\partial y} \frac{\partial K}{\partial y} \right). \quad (14)$$

If
$$z = x + iy, \quad (15)$$

then ψ_0 is the imaginary part of a function

$$w_0(z) = \phi_0 + i\psi_0, \quad (16)$$

which is the complex velocity-potential function for incompressible flow past the body with the undisturbed flow equal to unity.

To order M^2 it follows from (4) that

$$\frac{\rho_0}{\rho} = 1 + \frac{1}{2} M^2 (K - 1), \quad (17)$$

so
$$\frac{U}{U_0} = \frac{\partial \psi_0}{\partial y} + M^2 \left(\frac{K-1}{2} \frac{\partial \psi_0}{\partial y} + \frac{\partial \psi_1}{\partial y} \right), \quad (18)$$

$$\frac{V}{U_0} = -\frac{\partial \psi_0}{\partial x} - M^2 \left(\frac{K-1}{2} \frac{\partial \psi_0}{\partial x} + \frac{\partial \psi_1}{\partial x} \right). \quad (19)$$

The equations (14), (17) and (18) have been given by G. E. Nitzberg in an unpublished communication.

2. Solution for the Second Approximation to the Stream Function.

To solve the equation for ψ_1 we use the variables z and \bar{z} instead of x and y , where

$$\bar{z} = x - iy \quad (20)$$

and is the conjugate of z . (The same procedure has been adopted by

Imai and Aihara⁽⁴⁾ in solving for the second term in the expansion of the velocity potential *.) We have

$$\frac{dw_0}{dz} = \frac{\partial \psi_0}{\partial y} + i \frac{\partial \psi_0}{\partial x}, \quad (21)$$

so if \bar{w}_0 is the conjugate of w_0

$$K = \frac{dw_0}{dz} \frac{d\bar{w}_0}{d\bar{z}}, \quad (22)$$

and if Im denotes "the imaginary part of"

$$\begin{aligned} \frac{\partial \psi_0}{\partial x} \frac{\partial K}{\partial x} + \frac{\partial \psi_0}{\partial y} \frac{\partial K}{\partial y} &= \text{Im} \left\{ \left(\frac{\partial \psi_0}{\partial y} + i \frac{\partial \psi_0}{\partial x} \right) \left(\frac{\partial K}{\partial x} + i \frac{\partial K}{\partial y} \right) \right\} \\ &= \text{Im} \left\{ \frac{dw_0}{dz} \left(2 \frac{\partial K}{\partial \bar{z}} \right) \right\} \\ &= \text{Im} \left\{ 2 \left(\frac{dw_0}{dz} \right)^2 \frac{d^2 \bar{w}_0}{d\bar{z}^2} \right\}. \end{aligned} \quad (23)$$

$$\text{Hence} \quad \frac{\partial^2 \psi_1}{\partial z \partial \bar{z}} = -\frac{1}{4} \text{Im} \left\{ \left(\frac{dw_0}{dz} \right)^2 \frac{d^2 \bar{w}_0}{d\bar{z}^2} \right\}, \quad (24)$$

and the general solution for ψ_1 is

$$\psi_1 = \text{Im} \left\{ -\frac{1}{4} \frac{d\bar{w}_0}{d\bar{z}} \int \left(\frac{dw_0}{dz} \right)^2 dz + F_1(z) \right\}, \quad (25)$$

where F_1 is an arbitrary function. Now write

$$\frac{\partial \psi_0}{\partial y} = (1+u), \quad -\frac{\partial \psi_0}{\partial x} = v, \quad (26)$$

$$\text{so that} \quad \frac{dw_0}{dz} = 1+u-iv, \quad (27)$$

$$\text{and} \quad q^2 = K = \left| \frac{dw_0}{dz} \right|^2 = (1+u)^2 + v^2. \quad (28)$$

Since F_1 is an arbitrary function we may add any function of z we please to the expression in curly brackets in (25). We add

$$\frac{1}{4} \left(2 - \frac{dw_0}{dz} \right) \int \left(\frac{dw_0}{dz} \right)^2 dz, \quad (29)$$

$$\text{and, since} \quad 2 - \frac{dw_0}{dz} - \frac{d\bar{w}_0}{d\bar{z}} = -2u, \quad (30)$$

we may write

$$\psi_1 = \text{Im} \left\{ -\frac{1}{2} u \int \left(\frac{dw_0}{dz} \right)^2 dz + F(z) \right\}, \quad (31)$$

* A method of using the complex variable to evaluate the first-order compressibility correction to the velocity has lately been given by Kaplan⁽⁵⁾, but we did not find Kaplan's method convenient for our purpose.

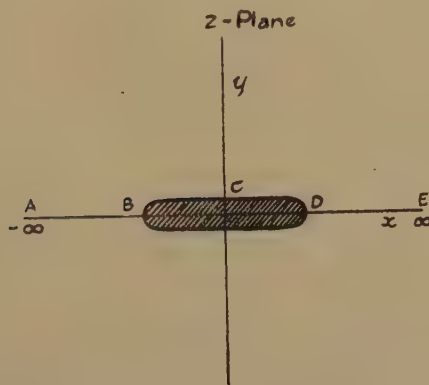
where $F(z)$ is a new arbitrary function. Further

$$\begin{aligned}\frac{\partial \psi_1}{\partial y} + i \frac{\partial \psi_1}{\partial x} &= -\frac{1}{2} \left(\frac{\partial u}{\partial y} + i \frac{\partial u}{\partial x} \right) \text{Im} \int \left(\frac{dw_0}{dz} \right)^2 dz - \frac{u}{2} \left(\frac{dw_0}{dz} \right)^2 + \frac{dF}{dz} \\ &= -\frac{i}{2} \frac{d^2 w_0}{dz^2} \text{Im} \int \left(\frac{dw_0}{dz} \right)^2 dz - \frac{u}{2} \left(\frac{dw_0}{dz} \right)^2 + \frac{dF}{dz}. \quad (32)\end{aligned}$$

3. Determination of the Complementary Function for Flow past a Symmetrical Body in Unlimited Fluid.

The arbitrary function $F(z)$ is now to be determined from the condition that ψ_1 vanishes at the solid boundaries. Hence the imaginary part of $F(z)$ is known at the boundaries, if, as we shall suppose, the irrotational incompressible flow is known. The value of $F(z)$ may therefore be determined by transforming the solid boundaries in the z -plane into a unit circle with its centre at the origin in a ζ -plane, and using Poisson's

Fig. 1 (a).



integral⁽⁶⁾. We are interested in flow past a symmetrical (uncambered) aerofoil or other obstacle at zero incidence, either in unlimited fluid or symmetrically placed in a channel. We shall consider first the necessary transformations in an unlimited fluid, but shall set them out in such a way that the extension to a channel of finite breadth is easy. On account of symmetry ψ_0 is zero not only on the surface of the body, but also on the axis of x if it is chosen to coincide with the axis of symmetry of the body, *i. e.* in fig. 1 (a) ψ_0 is zero along ABCDE; and because of symmetry it is sufficient if ψ_1 is also zero on ABCDE. Hence the imaginary part of $F(z)$ must cancel the other term in (31) along ABCDE.

Now write

$$w_0 = z + f(z), \quad (33)$$

so that

$$\begin{aligned}\text{Im} \int \left(\frac{dw_0}{dz} \right)^2 dz &= \text{Im} \left\{ z + 2f + \int \left(\frac{df}{dz} \right)^2 dz \right\} \\ &= 2\psi_0 - y + \text{Im} \int \left(\frac{df}{dz} \right)^2 dz. \quad (34)\end{aligned}$$

The integrals in (31), (32) and (34) may therefore be taken from $+\infty$ (*i. e.* from the point E) to the point z .

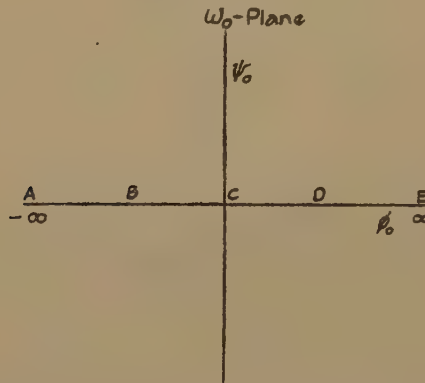
Also $(df/dz)^2$ has no singularities in the fluid and is certainly of a sufficiently small order at infinity to ensure that its integral round an infinite semi-circle vanishes, so

$$\int_{ABCDE} \left(\frac{df}{dz} \right)^2 dz = 0, \quad \dots \dots \dots (35)$$

and
$$\text{Im} \int_{ABCDE} \left(\frac{dw_0}{dz} \right)^2 dz = 0. \quad \dots \dots \dots (36)$$

Hence for a point on ABCDE the integral in (31) may be taken either from A or from E. On AB and DE, $v=0$, dw_0/dz is real, and the integral is purely real. Hence the imaginary part of $F(z)$ must vanish on AB and DE.

Fig. 1 (b).



On BCD let ϑ denote the angle which the direction of the resultant velocity, *i. e.* the tangent to the curve, makes with the axis of x . Then

$$\begin{aligned} \text{Im} \int \left(\frac{dw_0}{dz} \right)^2 dz &= \text{Im} \int K e^{-2i\vartheta} (ds e^{i\vartheta}) \\ &= - \int K \sin \vartheta ds = - \int K dy, \quad \dots \dots (37) \end{aligned}$$

and the integral may be taken either from B or from D, since the integral from B to D vanishes. Hence on BCD we must have

$$\text{Im} F(z) = -\frac{1}{2}u \int_0^y K dy = G, \text{ say.} \quad \dots \dots \dots (38)$$

The velocity potential ϕ_0 of the incompressible flow is a monotonic function, increasing from $-\infty$ to ∞ , on ABCDE, and G is a function of ϕ_0 on ABCDE, being zero on AB and DE.

In the w_0 -plane, ABCDE becomes the real axis (fig. 1 (b)), and the transformation

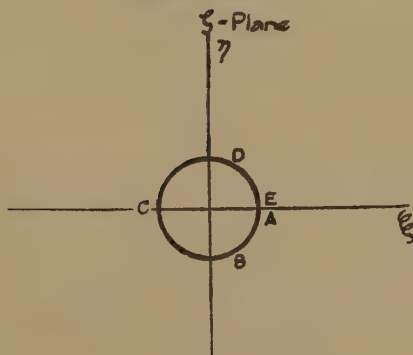
$$\zeta = \frac{w_0 + i}{w_0 - i} \quad \dots \dots \dots (39)$$

transforms the real axis in the w_0 -plane into the unit circle with its centre at the origin in the ζ -plane (fig. 1 (c)), the outside of the circle corresponding to the upper half of the w_0 -plane and to the upper half of the z -plane above ABCDE.

We shall require symbols for points on the boundaries as variables of integration in the ζ and w_0 -planes, and also symbols for general corresponding points outside or on the circle in the ζ -plane and above or on ABCDE in the w_0 and z -planes. We shall use ζ , w_0 and z for the latter; for the former we shall use t in the ζ -plane and w for the corresponding point in the w_0 -plane, so that the relation between t and w is the same as that between ζ and w_0 in (39). We shall still use G to denote the imaginary part of F on ABCDE in the z and w_0 -planes, and on the unit circle in the ζ -plane. Then, according to Poisson's Integral ⁽⁶⁾,

$$F(z) = \frac{1}{2\pi} \oint G \frac{\zeta + t}{\zeta - t} \frac{dt}{t} + \text{any real constant}, \quad (40)$$

Fig. 1 (c).



where the integral is taken in the positive sense round the unit circle in the ζ -plane. Since

$$\frac{\zeta + t}{\zeta - t} = i \frac{w_0 w + 1}{w_0 - w} \quad (41)$$

and

$$\frac{dt}{t} = -\frac{2i dw}{w^2 + 1}, \quad (42)$$

it follows that

$$F(z) = -\frac{1}{\pi} \int_{-\infty}^{\infty} G \frac{w_0 w + 1}{w_0 - w} \frac{dw}{w^2 + 1} + \text{any real constant}. \quad (43)$$

The integral in the w_0 -plane is taken along the real axis, so we may write $w = \phi$. Also G is zero on AB and DE, so if ϕ_1 and ϕ_2 are the values of ϕ_0 at B and D we may write

$$\begin{aligned} F(z) &= -\frac{1}{\pi} \int_{\phi_1}^{\phi_2} G(\phi) \frac{w_0 \phi + 1}{w_0 - \phi} \frac{d\phi}{\phi^2 + 1} + \text{any real constant} \\ &= -\frac{1}{\pi} \int_{\phi_1}^{\phi_2} G \frac{\phi d\phi}{\phi^2 + 1} + \frac{1}{\pi} \int_{\phi_1}^{\phi_2} \frac{G d\phi}{\phi - w_0} + \text{any real constant}. \end{aligned} \quad (44)$$

The first integral in the second line of equation (44) is real and independent of w_0 and z ; hence finally we may take

$$F(z) = \frac{1}{\pi} \int_{\phi_1}^{\phi_2} \frac{G(\phi) d\phi}{\phi - w_0}, \quad (45)$$

w_0 corresponding to z , and both having positive imaginary parts. When z tends to coincide with a point of the contour BCD, w_0 tends to a real value, say ϕ_0 , between ϕ_1 and ϕ_2 , and the integral in (45) is an improper integral if $w_0 = \phi_0$. To find the correct limit we therefore write

$$F(z) = \frac{1}{\pi} \int_{\phi_1}^{\phi_2} \frac{G(\phi_0) d\phi}{\phi - w_0} + \frac{1}{\pi} \int_{\phi_1}^{\phi_2} \frac{G(\phi) - G(\phi_0)}{\phi - w_0} d\phi. \quad . . . (46)$$

We have

$$\int_{\phi_1}^{\phi_2} \frac{d\phi}{\phi - w_0} = \log \left| \frac{\phi_2 - w_0}{\phi_1 - w_0} \right| + i \arg(\phi_2 - w_0) - i \arg(\phi_1 - w_0). \quad . (47)$$

In the limit, when $w_0 \rightarrow \phi_0$ and ϕ_0 is between ϕ_1 and ϕ_2 ,

$$\arg(\phi_2 - w_0) \rightarrow 0, \quad \arg(\phi_1 - w_0) \rightarrow -\pi, \quad (48)$$

since w_0 is in the upper half of the w_0 -plane. Hence

$$\int_{\phi_1}^{\phi_2} \frac{d\phi}{\phi - w_0} \rightarrow i\pi + \log \frac{\phi_2 - \phi_0}{\phi_0 - \phi_1}, \quad (49)$$

and when z is on the surface of the body

$$F(z) = iG(\phi_0) \frac{G(\phi_0)}{\pi} \log \frac{\phi_2 - \phi_0}{\phi_0 - \phi_1} + \frac{1}{\pi} \int_{\phi_1}^{\phi_2} \frac{G(\phi) - G(\phi_0)}{\phi - \phi_0} d\phi. \quad . (50)$$

For the derivative of F we have the equation

$$\frac{dF}{dz} = \frac{dw_0}{dz} \frac{dF}{dw_0} = \frac{dw_0/dz}{\pi} \int_{\phi_1}^{\phi_2} \frac{G(\phi) d\phi}{(\phi - w_0)^2}. \quad (51)$$

To find the limit when $w_0 \rightarrow \phi_0$ we write

$$\begin{aligned} \int_{\phi_1}^{\phi_2} \frac{G(\phi) d\phi}{(\phi - w_0)^2} &= \int_{\phi_1}^{\phi_2} \frac{G(\phi_0) d\phi}{(\phi - w_0)^2} + \int_{\phi_1}^{\phi_2} \frac{(\phi - \phi_0)G'(\phi_0)}{(\phi - w_0)^2} d\phi \\ &\quad + \int_{\phi_1}^{\phi_2} \frac{G(\phi) - G(\phi_0) - (\phi - \phi_0)G'(\phi_0)}{(\phi - w_0)^2} d\phi, \end{aligned} \quad (52)$$

where $G'(\phi)$ is $dG(\phi)/d\phi$, and we find that when z is on the surface of the body

$$\begin{aligned} \frac{dF}{dz} &= \frac{dw_0/dz}{\pi} \left\{ -\frac{(\phi_2 - \phi_1)G(\phi_0)}{(\phi_2 - \phi_0)(\phi_0 - \phi_1)} + G'(\phi_0) \left[i\pi + \log \frac{\phi_2 - \phi_0}{\phi_0 - \phi_1} \right] \right. \\ &\quad \left. + \int_{\phi_1}^{\phi_2} \frac{G(\phi) - G(\phi_0) - (\phi - \phi_0)G'(\phi_0)}{(\phi - \phi_0)^2} d\phi \right\}. \end{aligned} \quad (53)$$

4. Summary of Results. Calculation of Velocity on the Surface.

We may now summarize our results as follows:—The stream-function is given by equation (11); ψ_1 by equation (31) and $F(z)$ by equation (45), where G is given by (38). On the surface of the body $\psi_0 = 0$, and

$F(z)$ may be found either by evaluating the integral in (45) and then letting $w_0 \rightarrow \phi_0$, or numerically by the use of (46). For the velocity we have, from (18) and (19),

$$\frac{U-iV}{U_0} = \frac{dw_0}{dz} \left(1 + \frac{K-1}{2} M^2 \right) + M^2 \left(\frac{\partial \psi_1}{\partial y} + i \frac{\partial \psi_1}{\partial x} \right), \quad (54)$$

where K is given by equation (28) and $\partial \psi_1 / \partial y + i \partial \psi_1 / \partial x$ by equation (32); dF/dz is given in (51). On the surface of the body we may find dF/dz either by evaluating the integral in (51) (or evaluating the integral in (45) and differentiating with respect to w_0), or, if an analytical evaluation is impossible, numerically from equation (53).

One further point remains to be noted. At the surface of the body the velocity must be tangential to the surface, so $U-iV$ must be a real multiple of dw_0/dz . It follows from (54) that $\partial \psi_1 / \partial y + i \partial \psi_1 / \partial x$ must be a real multiple of dw_0/dz . This statement is verified in detail from equations (32) and (53) in the Appendix. It follows from (32) and (37) that at the surface of the body

$$\begin{aligned} \frac{\partial \psi_1 / \partial y + i \partial \psi_1 / \partial x}{dw_0/dz} &= \operatorname{Re} \left\{ \frac{i}{2} \frac{d^2 w_0 / dz^2}{dw_0 / dz} \int_0^y K dy - \frac{u}{2} \frac{dw_0}{dz} + \left(\frac{dF}{dw_0} \right)_{w_0=\phi_0} \right\} \\ &= -\frac{1}{2} \operatorname{Im} \left(\frac{d^2 w_0 / dz^2}{dw_0 / dz} \right) \int_0^y K dy - \frac{u}{2} - \frac{u^2}{2} + \operatorname{Re} \left(\frac{dF}{dw_0} \right)_{w_0=\phi_0}, \end{aligned} \quad (55)$$

so from (54) and (28)

$$\frac{U-iV}{U_0} = \frac{dw_0}{dz} + M^2 \frac{dw_0}{dz} \left\{ \frac{u}{2} + \frac{v^2}{2} - \frac{1}{2} \operatorname{Im} \left(\frac{d^2 w_0 / dz^2}{dw_0 / dz} \right) \int_0^y K dy + \operatorname{Re} \left(\frac{dF}{dw_0} \right)_{w_0=\phi_0} \right\}. \quad (56)$$

For the last term in (56) we have the equation

$$\begin{aligned} \operatorname{Re} \left(\frac{dF}{dw_0} \right)_{w_0=\phi_0} &= \frac{1}{\pi} \operatorname{Re} \lim_{w_0 \rightarrow \phi_0} \int_{\phi_1}^{\phi_2} \frac{G(\phi) d\phi}{(\phi - w_0)^2} \\ &= -\frac{(\phi_2 - \phi_1) G(\phi_0)}{(\phi_2 - \phi_0)(\phi_0 - \phi_1)} + G'(\phi_0) \log \frac{\phi_2 - \phi_0}{\phi_0 - \phi_1} \\ &\quad + \int_{\phi_1}^{\phi_2} \frac{G(\phi) - G(\phi_0) - (\phi - \phi_0) G'(\phi_0)}{(\phi - \phi_0)^2} d\phi. \end{aligned} \quad (57)$$

$G'(\phi_0)$ may be calculated as follows. Denote by ϑ , as before, the angle between the axis of x and the direction of the velocity, so that on the boundary ϑ is the angle between the axis of x and the tangent to the boundary curve. Let ds be an element of length of the boundary curve in the direction opposed to the velocity. Write

$$H = \int_0^y K dy, \quad G = -\frac{1}{2} u H \quad (58)$$

(from (38)). We have

$$\frac{dw_0}{dz} = 1 + u - iv = q e^{-i\vartheta}, \quad (59)$$

and
$$u = q \cos \vartheta - 1, \quad v = q \sin \vartheta. \quad (60)$$

Then
$$\frac{dH}{ds} = \frac{dH}{dy} \frac{dy}{ds} = -K \sin \vartheta, \quad (61)$$

and
$$\frac{dG}{ds} = \frac{1}{2} u K \sin \vartheta - \frac{H}{2} \left(\frac{dq}{ds} \cos \vartheta - \frac{q \sin \vartheta}{r} \right), \quad (62)$$

where r is the radius of curvature of the boundary curve, and is positive if ϑ decreases in the direction of the velocity, *i. e.* if the boundary curve is concave downwards above the axis of x . On the boundary

$$d\phi_0 = -q ds, \quad (63)$$

so
$$\frac{dG}{d\phi_0} = -\frac{1}{2} uv + \frac{H}{2} \left(\frac{\cos \vartheta}{q} \frac{dq}{ds} - \frac{\sin \vartheta}{r} \right). \quad (64)$$

Finally, on the boundary $dz = -e^{i\vartheta} ds$, so

$$\frac{d^2 w_0}{dz^2} = -\frac{d(qe^{-i\vartheta})}{e^{i\vartheta} ds} = -e^{-2i\vartheta} \left(\frac{dq}{ds} - \frac{iq}{r} \right), \quad (65)$$

and
$$\frac{d^2 w_0/dz^2}{dw_0/dz} = -e^{-i\vartheta} \left(\frac{1}{q} \frac{dq}{ds} - \frac{i}{r} \right), \quad (66)$$

$$\text{Im} \left(\frac{d^2 w_0/dz^2}{dw_0/dz} \right) = \frac{\sin \vartheta}{q} \frac{dq}{ds} + \frac{\cos \vartheta}{r}. \quad (67)$$

Our solution is now complete. In particular the velocity at any point of the surface is given by (56), (57) and (58); and (64) and (67) may be useful in carrying out the calculations. We shall illustrate the use of these equations by considering the simplest possible case, namely flow past a circular cylinder.

5. Results of a Circular Cylinder.

For a circular cylinder of radius a ,

$$w_0 = z + \frac{a^2}{z}, \quad \frac{dw_0}{dz} = 1 + u - iv = qe^{-i\vartheta} = 1 - \frac{a^2}{z^2},$$

and on the surface $z = ae^{i\vartheta}$,

$$\phi_0 = 2a \cos \theta, \quad u = -\cos 2\theta, \quad v = -\sin 2\theta,$$

$$q = 2 \sin \theta, \quad \vartheta = \theta - \frac{\pi}{2}, \quad \sin \vartheta = -\cos \theta,$$

$$\cos \vartheta = \sin \theta, \quad ds = a d\theta, \quad \frac{dq}{ds} = \frac{2}{a} \cos \theta,$$

$$r = a,$$

$$-\text{Im} \left(\frac{d^2 w_0/dz^2}{dw_0/dz} \right) = \frac{1}{a} \left(\frac{\cos^2 \theta}{\sin \theta} - \sin \theta \right) = \frac{\cos 2\theta}{a \sin \theta}$$

(from (67)).

$$K = q^2 = 4 \sin^2 \theta = \frac{4y^2}{a^2}; \quad H = \int_0^\vartheta K dy = \frac{4y^3}{3a^2} = \frac{4a}{3} \sin^3 \theta;$$

$$G = \frac{2a}{3} \sin^3 \theta \cos 2\theta.$$

$$F(z) = \frac{1}{\pi} \int_{\phi_1}^{\phi_2} \frac{G(\phi) d\phi}{\phi - w_0} = \frac{2a}{3\pi} \int_0^\pi \frac{\sin^4 \theta \cos 2\theta}{\cos \theta - W} d\theta,$$

where

$$W = w_0/2a.$$

Now

$$\begin{aligned} \sin^4 \theta \cos 2\theta &= 2 \cos^6 \theta - 5 \cos^4 \theta + 4 \cos^2 \theta - 1, \\ &= 2W^6 - 5W^4 + 4W^2 - 1 \\ &\quad + \{\cos \theta - W\} \left\{ \frac{1}{8} \cos 5\theta + \frac{W}{4} \cos 4\theta + \left(\frac{W^2}{2} - \frac{5}{8} \right) \cos 3\theta \right. \\ &\quad + \left(W^3 - \frac{3W}{2} \right) \cos 2\theta + \left(2W^4 - \frac{7W^2}{2} + \frac{3}{2} \right) \cos \theta \\ &\quad \left. + 2W^5 - 4W^3 + \frac{9W}{4} \right\}, \end{aligned}$$

and for $\text{Im}(W) > 0$,

$$\int_0^\pi \frac{d\theta}{\cos \theta - W} = \frac{\pi i}{(1 - W^2)^{1/2}}.$$

$$\text{Hence } F(z) = \frac{2a}{3} \left\{ 2W^5 - 4W^3 + \frac{9W}{4} + i(2W^2 - 1)(1 - W^2)^{3/2} \right\},$$

$$\text{and } \text{Re} \left(\frac{dF}{dw_0} \right)_{w_0 = \phi_0} = \frac{10}{3} \cos^4 \theta - 4 \cos^2 \theta + \frac{3}{4}.$$

Hence, from (56),

$$\begin{aligned} \frac{U - iV}{U_0} &= \frac{dw_0}{dz} \left\{ 1 + M^2 \left[-\frac{\cos 2\theta}{2} + \frac{\sin^2 2\theta}{2} + \frac{2}{3} \cos 2\theta \sin^2 \theta \right. \right. \\ &\quad \left. \left. + \frac{10}{3} \cos^4 \theta - 4 \cos^2 \theta + \frac{3}{4} \right] \right\}, \\ &= \frac{dw_0}{dz} \left\{ 1 + M^2 \left[\frac{7}{12} - \cos^2 \theta \right] \right\}, \end{aligned}$$

at the surface, and the factor agrees with that found by Janzen⁽¹⁾ and Rayleigh⁽²⁾.

6. Comparison with Linear Perturbation Theory for a Body of Small Thickness.

We may also show from (56) that the result there established is in agreement with the results of the linear perturbation theory⁽³⁾, according to which the excess velocity near the position of maximum thickness, for example, is increased by compressibility in the ratio $(1 - M^2)^{-1/2}$; for a body of small thickness t this result should be correct to order t . If we keep only terms in M^2 , the velocity at the position of maximum thickness should therefore be given by

$$\frac{U}{U_0} = 1 + \left(1 + \frac{1}{2} M^2 \right) u + O(t^2). \quad (68)$$

Now at the position of maximum thickness, $v = 0$, $\sin \vartheta = 0$, $\cos \vartheta = 1$; if the shape is kept constant apart from variations in thickness, the radius

of curvature at the position of maximum thickness varies as t^{-1} ; hence from (67)

$$\operatorname{Im}\left(\frac{d^2w_0/dz^2}{dw_0/dz}\right)=0(t). \quad (69)$$

Also, since the maximum value of K is of order unity

$$H=\int_0^y K dy=0(t). \quad (70)$$

Moreover u is of order t , and it follows from (56) that

$$\frac{U}{U_0}=1+\left(1+\frac{1}{2}M^2\right)u+M^2(1+u)\operatorname{Re}\left(\frac{dF}{dw_0}\right)_{w_0=\phi_0}+0(t^2), \quad (71)$$

where $\operatorname{Re}(dF/dw_0)_{w_0=\phi_0}$ is given by (57).

At the position of maximum thickness, in addition to $v=0$, $\sin \vartheta=0$, $u=0(t)$, we also have $dq/ds=0(t)$, and it follows from (58) and (64) that

$$G(\phi_0)=0(t^2), \quad G'(\phi_0)=0(t^2). \quad (72)$$

Remembering that ϕ_1 , ϕ_2 are the values at the trailing and leading edges, and ϕ_0 the value at the position of maximum thickness, we see that

$$-\frac{(\phi_2-\phi_1)G(\phi_0)}{(\phi_2-\phi_0)(\phi_0-\phi_1)}+G'(\phi_0)\log\frac{\phi_2-\phi_0}{\phi_0-\phi_1}=0(t^2). \quad (73)$$

We shall now show that

$$I=\int_{\phi_1}^{\phi_2}\frac{G(\phi)-G(\phi_0)-(\phi-\phi_0)G'(\phi_0)}{(\phi-\phi_0)^2}d\phi=0(t^2). \quad (74)$$

At the leading and trailing edges $u=-1$. Let ϕ_3 , ϕ_4 be the values of ϕ at the points, near the trailing and leading edges respectively, where $u=0-i$, *e. i.*, $q \cos \vartheta=1$. (We proceed on the assumption that the leading and trailing edges are stagnation points: if the trailing edge is cusped the only difference is that we do not need to consider the neighbourhood of the trailing edge separately.) Between ϕ_1 and ϕ_3 , and between ϕ_2 and ϕ_4 , H is $0(t)$ but u may be $0(1)$, so $G=0(t)$. But it is easily seen in simple examples, and may be proved generally for any thin streamline body from Theodorsen's theory⁽⁷⁾, that

$$\phi_3-\phi_1=0(t), \quad \phi_2-\phi_4=0(t). \quad (75)$$

The integrand in (74) is therefore $0(t)$ between ϕ_1 and ϕ_3 , and between ϕ_2 and ϕ_4 , and if J denotes the integrand,

$$\int_{\phi_1}^{\phi_3} J d\phi + \int_{\phi_4}^{\phi_2} J d\phi = 0(t^2). \quad (76)$$

Between ϕ_3 and ϕ_4 , u is $0(t)$ and G is $0(t^2)$. Also, between ϕ_3 and ϕ_4 we shall show that the integrand J is $0(t^2)$. If we keep at a distance from the position of maximum thickness equal to some finite fraction, say a quarter, of the chord, so that $\phi-\phi_0$ stays of order unity, we see

at once that J is $0(t^2)$, since $G(\phi)$, $G(\phi_0)$, $G'(\phi_0)$ are all $0(t^2)$. For a quarter of the chord on each side of the position of maximum thickness we set

$$J = \frac{1}{2} \frac{d^2 G}{d\phi^2}, \quad \dots \dots \dots (77)$$

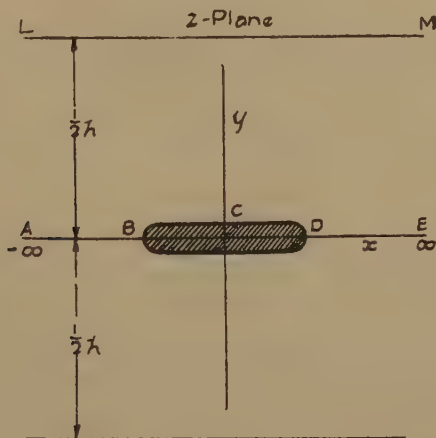
the value of $d^2 G/d\phi^2$ being taken at some intermediate point. Over the interval in question

$$\sin \vartheta, \quad dq/ds, \quad d^2 q/ds^2, \quad u, \quad v, \quad du/ds, \quad dv/ds, \quad 1/r, \quad H, \quad dH/ds$$

are all $0(t)$, and if we find $d^2 G/d\phi^2$ by using (64), with the help of (63) and (61), we see at once that $d^2 G/d\phi^2$ is $0(t^2)$ over the whole interval. Hence J is $0(t^2)$ over the interval (ϕ_3, ϕ_4) ,

and
$$\int_{\phi_3}^{\phi_4} J_1 d\phi = 0(t^2). \quad \dots \dots \dots (78)$$

Fig. 2 (a).



From (76) and (78) it follows that (74) is true; and from (73), (74) and (57) it follows that

$$\operatorname{Re} \left(\frac{dF}{dw} \right)_{w_0 = \phi_0} = 0(t^2), \quad \dots \dots \dots (79)$$

so (68) follows from (71), and we have shown that our results are in agreement with those of the linear perturbation theory.

7. Determination of the Complementary Function for Flow past a Body Symmetrically Placed in a Channel.

Let us now consider flow past a symmetrical body at zero incidence symmetrically situated in a channel of breadth h . (Fig. 2 (a)). Equations (18), (19), (31) and (32) hold as before; but we shall now find F in two parts, and shall denote the complete value by F_c , so that we must write F_c in place of F in (31) and (32). One part of F_c will be given by an integral along the wall LM , and we shall denote this part

by F_h ; the other part will be given by an integral along BCD, and we shall now use F to denote this part only. Consequently

$$F_c = F + F_h. \quad (80)$$

ψ_0 is zero on ABCDE and is $\frac{1}{2}h$ on LM. By symmetry it is sufficient if ψ_1 is zero on ABCDE and on LM, so we require

$$\text{Im} F_c(z) = \frac{1}{2}u \text{Im} \int \left(\frac{dw_0}{dz} \right)^2 dz = G, \text{ say,} \quad (81)$$

on ABCDE and on LM. If now we make the substitution (33), then

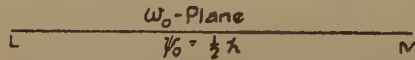
$$\text{Im} \int_{ABCDE} \left(\frac{df}{dz} \right)^2 dz = \text{Im} \int_{ALME} \left(\frac{df}{dz} \right)^2 dz = 0, \quad (82)$$

since the integrals along AL and ME are zero and that along LM is purely real. Hence

$$\text{Im} \int_{ABCDE} \left(\frac{dw_0}{dz} \right)^2 dz = 0. \quad (83)$$

We now fix the lower limit of the integral in (31), (32) and (81) as $z = \infty$, i.e., we take the integral from E. But because of (83) we obtain the

Fig. 2 (b).



same value if we take the integral from A, and it then follows, as before, that G vanishes on AB and DE. On BCD we have

$$G = -\frac{1}{2}u \int_0^v K dy = -\frac{1}{2}uH, \quad (84)$$

as before. Along LM the integral of $(dw_0/dz)^2$ is purely real, so for any point on LM

$$\text{Im} \int \left(\frac{dw_0}{dz} \right)^2 dz = \text{Im} \int_{EM} \left(\frac{dw_0}{dz} \right)^2 dz = \frac{h}{2}, \quad (85)$$

since u and v are zero on EM. Hence on LM,

$$G = \frac{1}{4}uh. \quad (86)$$

In the w_0 -plane, ABCDE becomes the real axis $\psi_0 = 0$, and LM the line $\psi_0 = \frac{1}{2}h$ (fig. 2 (b)). The transformation

$$Z = e^{2\pi w_0/h} \quad (87)$$

transforms the region between ABCDE and LM in the w_0 -plane into the upper half of the Z -plane, bounded by the real axis (fig. 2 (c)), and the transformation

$$\zeta = \frac{Z+i}{Z-i}, \quad (88)$$

then transforms the upper half of the Z -plane into the region outside

the unit circle with its centre at the origin in the ζ -plane (fig. 2 (d)). Then, as before,

$$F_c(z) = \frac{1}{2\pi} \oint G \frac{\zeta+t}{\zeta-t} \frac{dt}{t} + \text{any real constant}, \quad \dots \quad (89)$$

Fig. 2 (c).

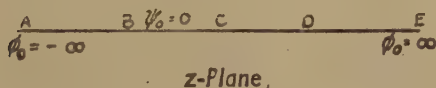
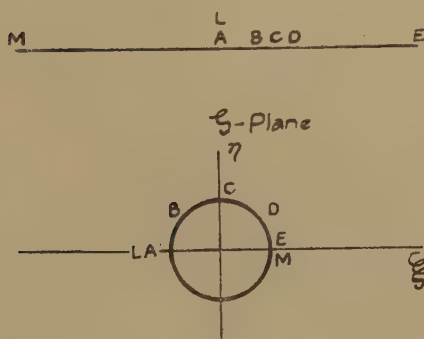


Fig. 2 (d).



where the integral is taken in the positive sense round the unit circle in the ζ -plane. Hence

$$\begin{aligned} F_c(z) &= \frac{1}{\pi} \oint_{EALME} G \frac{e^{2\pi(w+w_0)/h} + 1}{e^{2\pi w/h} - e^{2\pi w_0/h}} \frac{(2\pi/h) \cdot e^{2\pi w/h} dw}{e^{4\pi w/h} + 1} + \text{any real constant}, \\ &= \frac{1}{h} \oint_{EALME} G \frac{e^{4\pi w/h} dw}{e^{4\pi w/h} + 1} + \frac{1}{h} \oint_{EALME} G \frac{2e^{2\pi w/h} dw}{e^{2\pi w_0/h} - e^{2\pi w/h}} \\ &\quad + \text{any real constant}. \quad (90) \end{aligned}$$

The first integral vanishes, with G , on AL and ME , and it is purely real on EA and LM . Hence we may write

$$F_c(z) = \frac{1}{h} \oint_{EALME} G \left\{ \frac{e^{2\pi w_0/h} + e^{2\pi w/h}}{e^{2\pi w_0/h} - e^{2\pi w/h}} - 1 \right\} dw + \text{any real constant}, \quad (91)$$

and finally, since the integral of G is purely real,

$$F_c(z) = \frac{1}{h} \oint_{EALME} G \coth \frac{\pi}{h} (w_0 - w) dw. \quad \dots \quad (92)$$

Along AL and ME the integral vanishes. Also G vanishes on AB and DE . Hence the integral is to be taken along DCB , on which $w = \phi$, and along LM , on which $w = \phi + ih/2$. Hence if, as before, ϕ_1 and ϕ_2 are the values of ϕ_0 at B and D respectively,

$$F_c = F + F_h, \quad \dots \quad (93)$$

where

$$F = \frac{1}{h} \int_{\phi_1}^{\phi_2} G \coth \frac{\pi}{h} (\phi - w_0) d\phi$$

$$= -\frac{1}{2h} \int_{\phi_1}^{\phi_2} u H \coth \frac{\pi}{h} (\phi - w_0) d\phi, \quad (94)$$

and

$$F_h = -\frac{1}{h} \int_{-\infty}^{\infty} G \tanh \frac{\pi}{h} (\phi - w_0) d\phi$$

$$= -\frac{1}{4} \int_{-\infty}^{\infty} u \tanh \frac{\pi}{h} (\phi - w_0) d\phi, \quad (95)$$

the values of G and u to be taken being the values on BCD in (94) and on LM in (95).

$$\text{Also} \quad \frac{dF_c}{dw_0} = -\frac{\pi}{h^2} \oint_{\text{EALME}} G \operatorname{cosech}^2 \frac{\pi}{h} (w_0 - w) dw, \quad (96)$$

$$\frac{dF}{dw_0} = \frac{\pi}{h^2} \int_{\phi_1}^{\phi_2} G \operatorname{cosech}^2 \frac{\pi}{h} (\phi - w_0) d\phi$$

$$= -\frac{\pi}{2h^2} \int_{\phi_1}^{\phi_2} u H \operatorname{cosech}^2 \frac{\pi}{h} (\phi - w_0) d\phi, \quad (97)$$

and

$$\frac{dF_h}{dw_0} = \frac{\pi}{h^2} \int_{-\infty}^{\infty} G \operatorname{sech}^2 \frac{\pi}{h} (\phi - w_0) d\phi$$

$$= \frac{\pi}{4h} \int_{-\infty}^{\infty} u \operatorname{sech}^2 \frac{\pi}{h} (\phi - w_0) d\phi, \quad (98)$$

the values of G and u being again the values on BCD and on LM in (97) and (98), respectively.

When the point at which F_c or dF_c/dw_0 is being calculated tends to coincide with a point of the contour BCD, w_0 tends to a real value ϕ_0 between ϕ_1 and ϕ_2 . In the integrals for F_h and dF_h/dw_0 we may substitute ϕ_0 for w_0 , but to find the limiting values of F and dF/dw_0 we proceed in a similar fashion to that used in finding (50) and (53). The results we obtain are

$$(F)_{w_0=\phi_0} = i\Gamma(\phi_0) + \frac{G(\phi_0)}{\pi} \log \frac{\sinh [\pi(\phi_2 - \phi_0)/h]}{\sinh [\pi(\phi_0 - \phi_1)/h]}$$

$$+ \frac{1}{h} \int_{\phi_1}^{\phi_2} [G(\phi) - G(\phi_0)] \coth \frac{\pi}{h} (\phi - \phi_0) d\phi, \quad (99)$$

and

$$\operatorname{Re} \left(\frac{dF}{dw_0} \right)_{w_0=\phi_0} = -\frac{G(\phi_0)}{h} \left[\coth \frac{\pi}{h} (\phi_0 - \phi_1) + \coth \frac{\pi}{h} (\phi_2 - \phi_0) \right]$$

$$+ \frac{G'(\phi_0)}{\pi} \log \frac{\sinh [\pi(\phi_2 - \phi_0)/h]}{\sinh [\pi(\phi_0 - \phi_1)/h]}$$

$$+ \frac{\pi}{h^2} \int_{\phi_1}^{\phi_2} \left[G(\phi) - G(\phi_0) - \frac{hG'(\phi_0)}{2\pi} \sinh \frac{2\pi}{h} (\phi - \phi_0) \right] \operatorname{cosech}^2 \frac{\pi}{h} (\phi - \phi_0) d\phi.$$

(100)

8. *Velocity on the Surface of a Body Symmetrically Placed in a Channel.*

We now have all the required general formulæ. The stream-function is given by equation (11) and the velocity components by (18) and (19) or (54), and ψ_1 and its derivatives by (31) and (32) with F_c in place of F , where F_c is $F + F_h$ and the immediately preceding formulæ give the values of F and F_h and their derivatives. On the surface of the body, i. e., on BCD,

$$\frac{U - iV}{U_0} = \frac{dw_0}{dz} + M^2 \frac{dw_0}{dz} \left\{ \frac{u}{2} + \frac{v^2}{2} - \frac{1}{2} \operatorname{Im} \left(\frac{d^2 w_0 / dz^2}{dw_0 / dz} \right) \int_0^y K dy \right. \\ \left. + \operatorname{Re} \left(\frac{dF}{dw_0} \right)_{w_0 = \phi_0} + \left(\frac{dF_h}{dw_0} \right)_{w_0 = \phi_0} \right\}, \quad (101)$$

$$\text{where} \quad \left(\frac{dF_h}{dw_0} \right)_{w_0 = \phi_0} = \frac{\pi}{4h} \int_{-\infty}^{\infty} u \operatorname{sech}^2 \frac{\pi}{h} (\phi - \phi_0) d\phi, \quad (102)$$

the value of u being taken on the wall LM.

9. *Comparison with Linear Perturbation Theory for a Body of Small Thickness in a Channel.*

For a wide channel, if terms of order $1/h^4$ are neglected, $(dF_h/dw_0)_{w_0 = \phi_0}$ is fairly easily evaluated. If, with a suitable choice of origin, the complex velocity-potential function at a large distance from the body in an unlimited stream of unit velocity is

$$z + \frac{a_1}{z} + \frac{a^2}{z^2} + \dots, \quad (103)$$

then in the channel the value of u on LM is ⁽⁸⁾

$$u = \frac{\pi}{h} \left\{ a_1 \frac{d}{dx} - a_2 \frac{d^2}{dx^2} \right\} \tanh \frac{\pi x}{h} + O\left(\frac{1}{h^4}\right), \quad (104)$$

$$\text{and on LM,} \quad d\phi = (1 + u) dx \quad (105)$$

$$\phi = x + \text{constant} + \frac{\pi}{h} \left\{ a_1 - a_2 \frac{d}{dx} \right\} \tanh \frac{\pi x}{h} + O\left(\frac{1}{h^3}\right), \quad (106)$$

$$\text{so} \quad \operatorname{sech}^2 \frac{\pi}{h} (\phi - w_0) = \operatorname{sech}^2 \frac{\pi x}{h} + O\left(\frac{1}{h}\right), \quad (107)$$

the term in $1/h$ being odd in x . Hence

$$\left(\frac{dF_h}{dw_0} \right)_{w_0 = \phi_0} = \frac{\pi^2 a_1}{4h^2} \int_{-\infty}^{\infty} \operatorname{sech}^2 \frac{\pi x}{h} \frac{d}{dx} \left(\tanh \frac{\pi x}{h} \right) dx + O\left(\frac{1}{h^4}\right), \quad (108)$$

the terms of order $1/h^3$ in the integrand being odd in x and so giving no contribution to the integral. Hence

$$\left(\frac{dF_h}{dw_0} \right)_{w_0 = \phi_0} = \frac{\pi^2 a_1}{3h^2} = \lambda, \text{ say.} \quad (109)$$

Now in incompressible flow, if terms of order $1/h^4$ are neglected, the effect of the channel walls on the velocity at the surface of the body

is⁽⁸⁾ to increase it in the ratio $1+\lambda$. If, now, we consider the velocity at say, the position of maximum thickness, on a body of small thickness t , and denote by u_∞ the value of u in unlimited fluid, then, in a channel of breadth h ,

$$1+u=(1+\lambda)(1+u_\infty). \quad . \quad . \quad . \quad . \quad (110)$$

We may show, as before, that

$$-\frac{1}{2} \text{Im} \left(\frac{d^2 w_0 / dz^2}{dw_0 / dz} \right) \int_0^y K dy + \text{Re} \left(\frac{dF}{dw_0} \right)_{w_0=\phi_0} = 0(t^2). \quad . \quad . \quad (111)$$

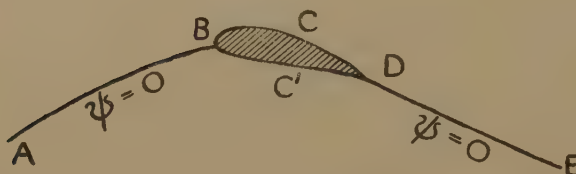
Hence

$$\begin{aligned} \frac{U-iV}{U_0} &= [1+\lambda][1+u_\infty] \left[1 + \frac{1}{2} M^2 (\lambda + \lambda u_\infty + u_\infty) + M^2 \lambda \right] + 0(t^2) \\ &= [1+u_\infty+\lambda] \left[1 + \frac{1}{2} M^2 (3\lambda + u_\infty) \right] + 0(t^2), \quad . \quad . \quad (112) \end{aligned}$$

since a_1 , and therefore λ , is $0(t)$. So finally

$$\begin{aligned} \frac{U-iV}{U_0} &= 1+u_\infty+\lambda + \frac{1}{2} M^2 u_\infty + \frac{3}{2} M^2 \lambda + 0(t^2) \\ &= 1+u + \frac{1}{2} M^2 u_\infty + \frac{3}{2} M^2 (u-u_\infty) + 0(t^2). \quad . \quad . \quad (113) \end{aligned}$$

Fig. 3.



The result in equation (113) is in agreement with the result of the linear perturbation theory, according to which u_∞ and $u-u_\infty$ are to be multiplied by $(1-M^2)^{-1/2}$ and $(1-M^2)^{-2/2}$, respectively, to allow for the effects of compressibility. To order M^2 these factors are

$$1 + \frac{1}{2} M^2 \quad \text{and} \quad 1 + \frac{3}{2} M^2,$$

respectively, as in (113).

10. Solution for Asymmetric Flow.

We may close this discussion with a few words about the flow past a body not symmetrical about the axis of x (fig. 3). Equations (31) and (32) still hold, and we have to determine F by the condition that on the surface of the body

$$\text{Im } F(z) = G = \frac{1}{2} u \text{Im} \int^z \left(\frac{dw_0}{dz} \right)^2 dz. \quad . \quad . \quad . \quad (114)$$

The streamline $\psi=0$ is no longer known, so we transform the whole boundary of the body, BCDC'B, into a unit circle. By the addition of

an arbitrary constant, if necessary, to w_0 , we make ϕ_0 have values equal in magnitude and opposite in sign, say a and $-a$, at D and B respectively. Then in the w_0 -plane BCDC'B transforms into both sides of the cut from $\phi=-a$ to $\phi=a$ along the real axis $\psi=0$, and the transformation

$$\frac{2w_0}{a} = \zeta + \frac{1}{\zeta}, \quad \dots \dots \dots (115)$$

transforms this cut into unit circle with its centre at the origin in the ζ -plane. The outside of the unit circle on the ζ -plane corresponds to the w_0 -plane outside the cut and to the z -plane outside BCDC'B. Then, as before,

$$F(z) = \frac{1}{2\pi} \oint G \frac{\zeta+t}{\zeta-t} \frac{dt}{t}, \quad \dots \dots \dots (116)$$

where the integral is taken in the positive sense round the unit circle in the ζ -plane. When we transform back to the w_0 -plane, we must note that G is to be considered as different functions of ϕ_0 , in general, on BCD and on BC'D. Let G be $G_1(\phi_0)$ on BCD and $G_2(\phi_0)$ on BC'D, and write

$$G_3 = \frac{1}{2}(G_1 + G_2), \quad G_4 = \frac{1}{2}(G_1 - G_2). \quad \dots \dots \dots (117)$$

Then we find that

$$F(z) = \frac{1}{\pi} \int_{-a}^a \frac{G_4(\phi) d\phi}{\phi - w_0} - \frac{i(w_0^2/a^2 - 1)^{1/2}}{\pi} \int_{-a}^a \frac{G_3(\phi) d\phi}{(\phi - w_0)(1 - \phi^2/a^2)^{1/2}}, \quad (118)$$

where that branch of $(w_0^2/a^2 - 1)^{1/2}$ is to be taken which is real and positive when w_0 is real and greater than a , and the value at other points is determined by the fact that the w_0 -plane is cut from $-a$ to $+a$. For a symmetrical body, G is an odd function of y , so that $G_2 = -G_1$, $G_3 = 0$, and on the upper surface $G = G_1 = G_4$; hence (118) agrees with our previous result (45).

Equation (118) exhibits the generalization of (45), but in considering examples it would often be simpler to transform the boundary of the body into a unit circle in the simplest possible way and use (116).

For the asymmetrical case of flow in a channel, the complete boundary in the z -plane transforms, in the w_0 -plane, into a segment of the real axis $\psi_0=0$ between two infinite lines $\psi_0=\text{constant}$ from $\phi_0=-\infty$ to $\phi_0=\infty$. The boundaries in the w_0 -plane may then be transformed into two concentric circles in the ζ -plane, and the value of F found from the analogue of Poisson's integral for the annulus so formed. The details are naturally much more complicated than in the cases considered previously.

The work described above was carried out in the Aerodynamics Department of the National Physical Laboratory, and it is published on the recommendation of the Aeronautical Research Committee and by permission of the Director of the Laboratory.

APPENDIX.

To prove that $\partial\psi_1/\partial y + i\partial_1/\partial x$ is a real multiple of dw_0/dz , it follows from (32) and (37) that we require to prove that

$$\operatorname{Im}\left\{\frac{iH}{2}\frac{d^2w_0/dz^2}{dw_0/dz} - \frac{u}{2}\frac{dw_0}{dz} + \left(\frac{dF}{dw_0}\right)_{w_0=\phi_0}\right\} = 0, \quad (1)$$

where

$$H = \int_0^y K dy$$

as in (58). From (48) and (64)

$$\operatorname{Im}\left(\frac{dF}{dw_0}\right)_{w_0=\phi_0} = \frac{dG}{d\phi_0} = -\frac{1}{2}uv + \frac{H}{2}\left(\frac{\cos\vartheta}{q}\frac{dq}{ds} - \frac{\sin\vartheta}{r}\right);$$

$$\text{from (66)} \quad \operatorname{Im}\left(\frac{iH}{2}\frac{d^2w_0/dz^2}{dw_0/dz}\right) = -\frac{H}{2}\left(\frac{\cos\vartheta}{q}\frac{dq}{ds} - \frac{\sin\vartheta}{r}\right),$$

$$\text{and clearly} \quad -\operatorname{Im}\left(\frac{u}{2}\frac{dw_0}{dz}\right) = \frac{1}{2}uv.$$

Hence (1) is satisfied.

References.

- (1) Janzen, *Phys. Zeit.* xiv. pp. 639–643 (1913).
- (2) Rayleigh, *Phil. Mag.* (6) xxxii. pp. 1–6 (1916); *Papers*, vi. pp. 402–408.
- (3) Goldstein and Young, "The Linear Perturbation Theory of Compressible Flow with Application to Wind Tunnel Interference," Aeronautical Research Committee, Reports and Memoranda No. 1909 (1943).
- (4) Imai and Aihara, "On the Subsonic Flow of a Compressible Fluid past an Elliptic Cylinder," *Aero. Res. Inst., Tokyo Imperial University*, Report No. 194 (1940).
- (5) Kaplan, "On the use of Residue Theory for Treating the Subsonic Flow of a Compressible Fluid," National Advisory Committee for Aeronautics (U.S.A.), Report No. 728 (1942).
- (6) Titchmarsh, 'The Theory of Functions,' pp. 124, 125. (Oxford, 1932.)
- (7) Theodorsen, "Theory of Wing Sections of Arbitrary Shape," National Advisory Committee for Aeronautics (U.S.A.) Report No. 411 (1931).
Theodorsen and Garrick, "General Potential Theory of Arbitrary Wing Sections," National Advisory Committee for Aeronautics (U.S.A.), Report No. 452 (1933).
- (8) Goldstein, "Two-Dimensional Wind Tunnel Interference," Aeronautical Research Committee, Reports and Memoranda, No. 1902 (1942).

LXIV. Notices respecting New Books.

A Shorter History of Science. By Sir WILLIAM CECIL DAMPIER (formerly WHETHAM), Sc.D., F.R.S. [Pp. x+189.] (Cambridge: University Press, 1944. Price 7s. 6d. net.)

SIR WILLIAM DAMPIER's larger volume, 'A History of Science and its Relations with Philosophy and Religion,' is deservedly well known, and three editions (one published during the war) have been exhausted. In this much smaller volume, the author re-focuses his outlook for the benefit of general readers who find the philosophic parts of the larger book difficult; and, without ignoring altogether the connexion between science and other activities, he aims at what he calls "a straightforward story of the growth of science reduced

to its simplest terms." A no less important aim that the author has in mind is to enable older schoolboys to achieve a humanist outlook with regard to science; indeed, it seems vastly important to ensure that the generation that is growing up during a period of what is called "scientific" war may get the right and not a wrong view of science and its social consequences (and abuses), a need perhaps even more urgent in education generally than in scientific training in particular.

With such aims before him, Sir William describes the rise of science from its remote beginnings, through the Greek and Roman periods, the Middle Ages and the Renaissance, to the first physical synthesis under Galileo and Newton. He then passes to an interesting chapter on the eighteenth century, followed by chapters on the physics, chemistry and biology of the nineteenth century, more recent biology, the new physics and chemistry and the stellar universe.

There is a useful "Note on Bibliography," in which the journal published in this country dealing with the history of science ('Annals of Science') might have been included.

There are 9 Plates and 14 text-figures: by war-time standards the production is good, but the price is high. D. McKIE.

Berkeley's Philosophical Commentaries. Edited by Prof. A. A. LUCE, Litt.D. [Pp. xlii+486.] (Edinburgh: Thomas Nelson & Sons, Ltd., 1944. Price £3 13s. 6d. net.)

WHATEVER our views on his philosophy, we must agree that George Berkeley, Bishop of Cloyne, was an extraordinary man. By 1710, when only 25 years old, he had written two small mathematical books, his 'New Theory of Vision' and the 'Treatise concerning the Principles of Human Knowledge,' which contained his famous philosophical system. He followed this, three years later, by 'Three Dialogues,' in which the same system was further expounded, but not further developed. Indeed, there was no further development: as Sir Alexander Lindsay has remarked: "Berkeley's thought matured early, but developed little. His philosophical writings are a reiteration . . . of the one central principle, which, as his *Commonplace Book* has shown us, struck upon his mind with convincing force when he was only twenty-one."

The volume before us contains an elaborate reprint of this *Commonplace Book*, with an introduction and copious notes by Prof. Luce. The *Book* itself raises questions rather than answers them; it contains 888 random jottings, reflections and queries, intended apparently to jog the author's memory at some later stage, and perhaps provide material for incorporation in a book. But we cannot study the germination of his philosophical ideas here, for they appear already in their full maturity. We find, for instance:—

"(472) You ask me whether the books are in the study now when no one is there to see them. I answer yes. you ask me are we not in the wrong for imagining things to exist when they are not actually perceiv'd by the senses. I answer no. the existence of our ideas consists in being perceiv'd, imagin'd thought on whenever they are imagin'd or thought on they do exist. Whenever they are mention'd or discours'd of they are imagin'd & thought on therefore you can at no time ask me whether they exist or no, but by reason of y^t very question they must necessarily exist."

"(98) The trees are in the park, that is, whether I will or no whether I imagine anything about them or no, let me but go thither & open my eyes by day & I shall not avoid seeing them."

But here, as in his later writings, we feel there is something lacking in directness, in force, and almost in candour. Berkeley always manages to shy away from the main point; what the plain man wants to know is why the trees "continue to grow, when there's no one about in the quad." Berkeley

wrapped up his answer in such subtle and sophisticated modes of diction, and interlarded it with so many fallacies repugnant to science or to common sense or to both, that it was hard for anyone to take it at its simple face-value. The plain man's reaction was typified by Dr. Johnson: "After we came out of church, we stood talking for some time of Bishop Berkeley's ingenious sophistry to prove the non-existence of matter, and that everything in the world is merely ideal . . . I shall never forget the alacrity with which Johnson answered, striking his foot with mighty force against a large stone, till he rebounded from it, 'I refute it thus.'" (Boswell's 'Life.')

A number of the entries are of scientific interest, as, for instance :—

"(303) Qu : why we see objects greater in ye dusk whether this can be solv'd by any but my principles."

"(450) Motion on 2^d thoughts seems to be a simple idea."

"(450 a) Motion distinct from ye thing moved is not conceivable."

"(451) Mem : To take notice of Newton for defining it also of Locke's wisdom in leaving it undefin'd."

"(876) If there were only one ball in the world it could not be moved. there could be no variety of appearance."

Such entries show that the future Bishop had a keen eye for the philosophical problems of science, but neither here nor in his subsequent writings does he show any capacity for solving them. In some ways the physics of to-day is nearer to Berkeley's philosophy than the physics of his own day was, or could possibly have been, but Berkeley shows no realization that such an approach could possibly occur, and so of course no anticipation of the mentalist tendencies which many find in modern physics.

The interest of the book is interrupted at frequent intervals by venomous and entirely irrelevant attacks on mathematicians. Dr. Luce suggests that this is because the mathematicians that Berkeley read were "more or less identified with one or other of certain tenets to which he strongly objected—materialism, extensionism, abstract ideas, infinite divisibility, and irreligion." This may have been especially the case with infinite divisibility; the idea always made him bristle. In the 'Principles' he assures us that there is no such thing as the 10,000th part of an inch; *esse est percipi*, so that if a line is too short to be perceived, it cannot exist. We might suspect that Berkeley had never heard of a microscope, but he had. ('New Theory of Vision,' cxxxii. 1710 edition.)

Even so, Dr. Luce's suggestion does not explain Berkeley's colossal misunderstandings of mathematics. He asserts that he can square the circle, while other mathematicians cannot (395), that particular circles can be squared (249), but apparently not all, that there is no such constant as π , since the circumference and diameter "are not proportional in all circles; hence tis nonsense to seek the terms of one general proposition whereby . . . to square all circles" (457). He also maintains (470) that "a mean proportional cannot be found betwixt any two given lines," and (500) that "one square cannot be double of another. Hence the Pythagoric Theorem is false." His general verdict on mathematicians is that he can "see no wit in any of them but Newton" (372), while "Newton's harangue amounts to no more than that gravity is proportional to gravity" (361). Prof. Luce's notes bear evidence of great scholarship and research, but they do not conceal that Berkeley had a one-track mind which led always to the same single, although perhaps superlatively good, idea.

The volume is sumptuously produced on 528 pages of superfine quarto paper, many dozens of which are entirely blank. Some readers may wonder if this was "really necessary" in these days of paper stringency, and so, perhaps, may some purchasers when they notice the price of the book.

J. H. J.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

LXV. *On the Tones of Thin Cylindrical Tubes.*

By GEORGE E. ALLAN, D.Sc.*

[Received June 20, 1944.]

Introductory.

A THIN cylindrical tube, when struck or otherwise caused to sound, vibrates with a frequency which depends on the material of which the tube is made, the dimensions and uniformity of its cylindrical shape, and the manner in which the tube is supported and is struck.

Lateral or Transverse Vibrations.

If a tube is held in a vertical position between thumb and forefinger, at a distance of about two-ninths (0.2242) of its length from the upper end, and struck near the middle or end, stationary waves are set up in the tube about two nodes equidistant from the ends, and the tube executes transverse vibrations about these nodes in much the same way as a bar on a harmonicon or xylophone vibrates.

Peripheral Vibrations.

If the tube is held at one end and struck at the other end, the chief tone elicited is caused by vibrations of the portion of the tube round the free edge, and the vibrations resemble those of a sounding bell. The stationary waves set up round the tube are constant in frequency for different lengths of the same tube, and the tone heard may, as in the first case, be accompanied by overtones.

Longitudinal Vibrations.

In addition to the above, the tube is capable of emitting tones due to longitudinal vibrations not involving lateral displacement. For example, when it is clamped or held at the middle and struck at the end in the direction of its length or stroked lengthwise with a prepared cloth, stationary waves of high frequency are set up.

In the present paper these three modes of vibration and the relation to each other of the tones they produce are investigated at ordinary temperature and pressure, chiefly with the aid of thin brass tubes.

Theory.

Lord Rayleigh has given in his first volume on sound ⁽¹⁾ the theory of lateral vibrations of thin straight elastic rods or bars. The frequency

* Communicated by the Author.

of these lateral vibrations is derived in terms of the velocity of transmission of longitudinal waves along the bar, which is not under any externally applied tension. The theory is also developed by Donkin⁽²⁾ and Violle⁽³⁾. Chree has calculated⁽⁴⁾ an expression for the frequency of longitudinal vibrations to a second degree of approximation, for straight rods of circular, elliptical or rectangular cross-section. Kalähne has investigated⁽⁵⁾ the effect of the air or other gas which jackets a tube inside and out and acts as an additional mass in causing a slight lowering of the frequency of the tube.

Transverse Axial Vibrations.

Rayleigh gives the frequency of a bar or rod as $kVm^2/2\pi L^2$, where k is the radius of gyration of the section of the vibrator about an axis through its centroid, V the velocity of longitudinal waves along the bar, m a constant for a given partial, and L the length of the bar.

Assuming that the formula may be applied to thin tubes of small diameter compared with the length, we have, for a tube vibrating transversely, $k^2 = (R^2 + r^2)/4$, where R and r are the radii of the circular section. Using the corresponding diameters, $k^2 = (D^2 + d^2)/16$. Let D_m , the mean diameter, be taken as $\sqrt{(D^2 + d^2)}/2$ for the thin tubes considered, then $k^2 = D_m^2/8$, and the frequency N of the fundamental tone of the tube is $(mVD_m)/4\sqrt{2}\pi L^2$. For the fundamental, m is 4.7300, and the expression for the velocity becomes

$$V = NL^2/1.259 D_m \quad \dots \quad (1)$$

It follows that with tubes of equal diameter the frequency is inversely proportional to the square of the length, and for tubes of equal length the frequency is directly proportional to the mean diameter. Since the frequency of the fundamental of a tube is measurable with simple means within a certain range, it is possible to test the last formula by experiment.

Experimental.

The product NL^2 .

After preliminary observations with tubes of glass, iron and brass, the constancy of the product NL^2 was tested in the following way. A long thin brass tube, rather more than an inch in diameter was obtained, its weight, length and diameters were measured, the nodal positions for the fundamental calculated and marked, and the tube placed on rubber rests at these nodes and the frequency of the fundamental determined by monochord⁽⁶⁾. Then an inch or thereabouts was cut off from one end, the end squared, the tube wiped within and without, and the above measurements repeated. In this way a long series of readings for two tubes was taken and tabulated. From these readings the variation of frequency with length could be determined. The length varied from 184 to 25 cm., and the frequency from 40 to 1900. The quantities tabulated are: number of tube, length, frequency, NL^2 , weight, weight

per cm., length/mean diameter. The results are given in three tables in order of diminishing length of tube. The number assigned to each tube is chronological, and the figures in Table III. were those first obtained. In Table I. the long tubes had low frequencies which were difficult to assess accurately. Accordingly the frequency N_5 of the fifth partial of the tube was measured and that of the fundamental calculated from the relation $N=N_5/12.19$, an average found by measurements of the partials of different tubes (see p. 580). The ratio of the weight of the tube to its length increased as portions were cut off from the same end, as far as tube No. 69 (Table I.). With the remainder the shortening

TABLE I.

Brass tubes with frequency range 40 to 107. Mean diameter 2.777 cm.
Thickness 0.077 cm.

No.	L cm.	N.	NL^2 $\div 10^5$.	W gm.	W/L gm./cm.	L/D _m .
49	183.7	40.5	13.74	1074.7	5.850	67.48
51	175.9	44.1	13.69	1028.9	5.850	63.40
53	169.9	47.2	13.66	994.9	5.855	61.22
55	164.0	50.6	13.65	960.95	5.858	59.07
57	158.8	53.8	13.59	931.2	5.863	56.28
59	153.7	57.3	13.56	901.5	5.865	55.35
61	148.7	61.1	13.52	872.6	5.868	53.55
63	143.6	65.3	13.46	843.1	5.871	51.71
65	138.4	70.1	13.42	813.3	5.875	49.85
67	133.2	75.5	13.37	782.8	5.879	47.95
69	128.0	81.3	13.30	753.1	5.883	46.10
71	123.0	87.9	13.28	723.0	5.879	44.28
73	117.8	95.2	13.20	692.4	5.876	42.43
75	112.7	103.1	13.09	661.9	5.874	40.58
76	110.8	107.0	13.13	650.5	5.872	39.89

took place at the other end, and the ratio W/L diminished. The product NL^2 was corrected in all cases to a uniform value of $W/L=5.873$ gm./cm., and NL^2 is not constant but diminishes as the tube is shortened.

For convenience in tabulation the values of NL^2 are divided by 10^5 . Also every alternate point has been omitted except the last, as the frequencies were quite close together. All the points are included on the graph.

In Table II. the frequencies were determined, as before, from the fifth partials for the tubes Nos. 77 to 86; thereafter the monochord was tuned to the fundamental of the tube, comparison of the results by the two methods having shown satisfactory agreement. In this series the values of W/L are nearly constant, because the tube was shortened alternately from opposite ends. As before, the product NL^2 diminishes slightly as the tube is shortened.

TABLE II.

Same tube. Range of frequencies from 111 to 339.

No.	L cm.	N.	NL ² ÷ 10 ⁵ .	W gm.	W/L gm./cm.	L/D _m .
77	108.8	110.8	13.14	639.2	5.873	39.19
78	107.1	114.2	13.10	629.4	5.874	38.58
79	105.6	117.3	13.08	620.2	5.873	38.03
80	104.2	120.3	13.07	612.0	5.873	37.52
81	102.9	122.9	13.01	604.0	5.872	37.04
82	100.9	127.3	12.95	592.7	5.875	36.33
83	98.41	133.5	12.93	578.0	5.874	35.46
84	95.73	140.6	12.88	562.4	5.875	34.50
85	93.22	147.8	12.84	547.5	5.873	33.59
86	90.70	155.7	12.81	532.8	5.874	32.69
87	88.16	163.6	12.72	517.8	5.873	31.78
88	85.67	173.5	12.73	503.2	5.874	30.87
89	82.98	185.3	12.76	487.3	5.872	29.90
90	80.40	197.7	12.78	472.3	5.874	28.97
91	77.79	210.1	12.71	456.8	5.872	28.03
92	75.39	224.1	12.74	442.8	5.873	27.17
93	70.25	257.1	12.69	412.3	5.870	25.32
94	65.19	297.8	12.65	382.9	5.874	23.49
95	61.10	338.9	12.63	335.0	5.873	20.56

TABLE III.

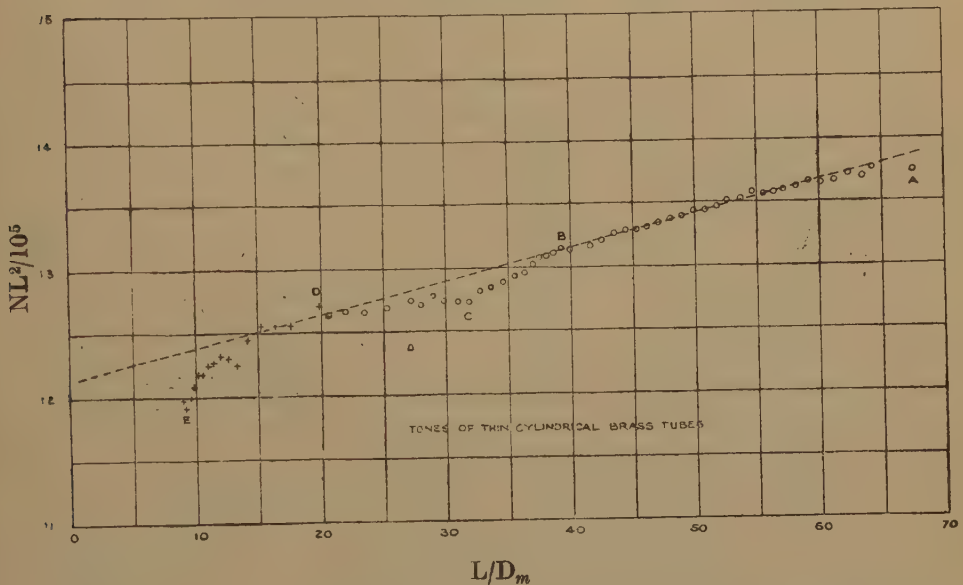
Second tube. Frequency range 427 to 1908. Mean diameter 2.791 cm.
Thickness 0.079 cm.

No.	L cm.	N.	NL ² ÷ 10 ⁵ .	W gm.	W/L gm./cm.	L/D _m .
32	54.68	427	12.72	—	—	19.60
33	49.08	523	12.56	290.85	5.926	17.59
34	45.77	601	12.56	271.05	5.923	16.41
35	42.48	696	12.53	251.42	5.919	15.23
36	39.73	789	12.44	235.05	5.916	14.24
37	37.12	889	12.24	219.47	5.912	13.30
38	35.09	999	12.30	207.24	5.906	12.58
39	33.50	1097	12.32	197.80	5.904	12.01
40	32.02	1195	12.27	188.89	5.899	11.48
41	30.70	1296	12.24	181.06	5.896	11.00
42	29.56	1392	12.18	174.33	5.897	10.59
43	28.51	1496	12.18	168.08	5.895	10.22
44	27.59	1584	12.08	162.64	5.894	9.89
45	26.70	1678	11.99	157.42	5.895	9.57
46	25.76	1790	11.91	151.76	5.892	9.23
47	25.16	1888	11.98	148.21	5.891	9.02
48	25.00	1908	11.96	147.27	5.891	8.96

In Table III. are collected the first measurements made by the method described. The tube in this case was slightly wider than the first one, and the value of W/L is a little greater than in Tables I. and II. The product NL^2 was corrected to the same value as before, and the material which the table offered was considered suitable for inclusion in the graph. As in the previous two series the product NL^2 diminishes as L is made shorter, and it can be seen that whether the ratio of weight of tube to length increases, is constant or decreases the product NL^2 diminishes.

A graph is given in which $NL^2/10^5$ is plotted against L/D_m , the figures of Table III. being included. The portion AB was obtained from Table I., the part BCD from Table II., and from D to E from Table III. The straight line shown on the graph has the equation :

$$NL^2 = 12.13 \times 10^5 + 2570 L/D_m. \quad . \quad . \quad . \quad . \quad . \quad (2)$$



For small values of length/diameter the graph falls away. This occurs in the region in which the frequency of the transverse vibrations coincides with the frequency of the peripheral vibrations, namely for a value of L/D_m in this case about 9.

The ratio N/D_m .

In testing the constancy of the ratio of the fundamental frequency of the tube to its mean diameter, no special tubes were employed. Tubes of different lengths and diameters, already tested, were chosen and arranged in order of increasing diameter. The value of NL^2 was calculated or known for each tube, and, making use of the graph resulting from the first tests, a straight-line law was assumed as the relation between NL^2 and L/D_m , and that, for thin brass tubes of different

diameters, the same relation could be represented by straight lines parallel to the original line. The frequencies of tubes of different diameters and 50 cm. in length could then be approximately estimated by a graphical method. For example, taking the figures in Table IV. for tube No. 14, the point $NL^2/10^5=5.47$, $L/D_m=51.7$ was marked and a straight line drawn through it parallel to the original line. The abscissa 42.4 ($=50/D_m$) had a corresponding ordinate 5.23 ($=NL^2/10^5$). From the equation $N_{50}50^2/10^5=5.23$, $N_{50}=209.2$, and similarly with the other figures. The results are given in the table.

The mean value of N_{50}/D_m is 179.7, and the result may be expressed in the form

$$N_{50}=180 D_m, \quad \dots \dots \dots (3)$$

or, a thin brass tube 50 cm. long has, when sounding its fundamental tone, a frequency of lateral vibrations 180 times the mean diameter of

TABLE IV.

Frequencies of tubes 50 cm. long. Mean diameters from 1.18 to 3.63 cm.

Tube no.	Length cm.	N.	D_m cm.	L/D_m .	$NL^2 \div 10^5$.	$50/D_m$.	N_{50} .	N_{50}/D_m .
14	61.0	147	1.18	51.7	5.47	42.4	209.2	177.2
10	39.1	445	1.53	25.6	6.81	32.7	278.0	181.7
9	82.16	147	2.13	38.6	9.92	23.5	380.4	178.6
18	50.0	485	2.72	18.4	12.13	18.4	485.0	178.2
96	57.04	388	2.77	20.6	12.63	18.05	501.6	181.0
22	50.0	545	3.01	16.62	13.63	16.62	545.0	181.0
—*	60.56	447.4	3.63	16.68	16.41	13.77	653.2	179.9

* A tube used in the gong experiments, see p. 581.

the tube. If the material is other than brass, the constant 180 is multiplied by the ratio of the velocities of sound in the two materials, as shown, for example, in a succeeding table.

The Frequency Formula as Verified by Experiment.

When the figures for length of tube, mean diameter and frequency are plotted in the form $\log L/\log N/D_m$, the result is a straight line with intercepts on the positive parts of the axes, and the equation to the line may be expressed in the form

$$x \cos \theta + y \sin \theta = p, \quad \text{or} \quad \log \frac{N}{D_m} \cos \theta + \log L \sin \theta = p.$$

This, again, may be written in the more useful form,

$$\log \frac{N}{D_m} + \log L \tan \theta = p \sec \theta.$$

Measurements of the inclination of the perpendicular p to the axis of $\log N/D_m$ gave an average angle of $63^\circ 47'$, of which the tangent is 2.031. The value 2 was adopted for the tangent and the equation becomes

$$\log \frac{N}{D_m} + 2 \log L = p \sec \theta, \quad (4)$$

which is suitable for calculating the frequency of the lateral vibrations of a brass tube for which L/D_m is not too small, and for which $p \sec \theta = 5.661$. It may also assume the form

$$\frac{NL^2}{D_m} = 10^{p \sec \theta}.$$

When results for copper, steel, glass and aluminium were plotted on the same graph, these formed straight lines parallel to that for brass, their distances from the origin increasing with the velocity of sound, so

TABLE V.

Velocity of sound in different materials.

Material.	V metre/sec.	V/V ₁ .
Brass (70 p.c.Cu) . . .	3647 = V ₁	1.0
Zinc	3670	1.007
Copper	3846	1.054
Steel	5112	1.398
Aluminium	5123	1.405
Glass	5410	1.484

that by adding the term $\log V/V_1$ to the right side of equation (4) the equation may be made to apply generally, namely :

$$\log \frac{N}{D_m} + 2 \log L = p \sec \theta + \log \frac{V}{V_1} \quad (5)$$

The above table contains the results for the velocity of sound in some materials and also the ratio V/V_1 .

Equation (5) can be put into the form

$$\frac{NL^2}{D_m} = \frac{V}{V_1} 10^{p \sec \theta}. \quad (5a)$$

Taking, for example, the value 5.661 for brass, and $V/V_1 = 1$, the right side = 458,100, and this figure when divided by 364,700 (velocity of sound in brass in cm./sec.) gives the constant 1.256, from which it can be seen that the above equation is reducible to almost the same as the theoretical equation (1).

The extent of the agreement of the frequency found from the mono-

chord and that calculated from the $p \sec \theta$ equation may be gauged from the following table :—

TABLE VI.

Frequency by monochord and by calculation.

Tube no.	N monoc.	N calc.
49	40.5	37.7
54	48.8	45.7
59	57.3	53.9
64	67.6	65.3
69	81.3	77.7
74	99.1	95.7
79	117.3	114.1
84	140.6	138.9
89	185.3	184.8
94	297.8	299.3
34	601.0	610.3
39	1097	1140
44	1584	1681
48	1908	2046

From the tables and graph the value of NL^2 for a thin brass tube diminishes irregularly with L . When the tube is long, NL^2 may be obtained approximately from a straight-line equation such as the one given, but as the length decreases the effect of other vibrations which are present is shown by the departure of the graph from rectilinearity. Savart studied the coexistence of longitudinal and lateral vibrations in a horizontal bar and found that the nodes were much more numerous than those proper to the longitudinal vibrations alone. He concluded that the nodes were due to a transverse vibration synchronous with the longitudinal one, and stated the conditions under which synchronism existed. This coupling of the two modes of vibration no doubt exists also in a vibrating tube, and its presence is evident from the course of the graph and also from the fact that the series of partials for the peripheral tones follows the same law of relative magnitude as the series of partials for transverse tones. The irregularities in the graph render it impossible to apply a straight-line law for small values of L/D_m ; also the factor 1.259 in the simple theoretical formula for V may be regarded as an approximation applicable to brass tubes for which L/D_m is, say, greater than 35 or thereabouts.

The Peripheral Tone Predominant.

It is probable that for brass tube lengths such that L/D_m is less than 9 the peripheral tone becomes predominant and persists as the length of the short tube is further reduced. The tone which is now loudest is

that which conforms to the frequency law derived by Rayleigh and verified experimentally by Fenkner, namely, $N=kt/D_m^2$, where t is the thickness of the tube and k a constant with the value 188100 for brass. A number of rings of various sizes was tested by suspending them in turn from a kind of swing formed of stout thread. In all cases the frequency of the note emitted by the ring when allowed to hit the mouth of a binaural stethoscope was in rough agreement with that calculated from the above formula. Rayleigh has derived an expression for the frequency of a ring which should be lower in pitch than the peripheral tone, but no tone distinct from the peripheral one was observed.

Relation between the Frequencies of the Longitudinal and Transverse Vibrations of a Thin Cylindrical Tube.

If the frequency of the fundamental tone of a rod, bar or tube clamped at the middle and vibrating longitudinally is N_L , and that of the rod, etc., when vibrating transversely between two nodal rests is N_T , Violle has given the following values for the ratio N_L/N_T , namely,

Rod of circular section $N_L/N_T=0.5618$ length/diameter.

Bar of rectangular section $N_L/N_T=0.4863$ length/depth.

It will be shown in what follows that the ratio for a thin tube has the approximate value

$$N_L/N_T=0.4 \text{ length/mean diameter.} \quad (6)$$

Tubes of brass, copper and monel metal were measured to obtain the quantities N_L , N_T , L and D_m , with the results as shown in Table VII.

TABLE VII.

Relation of longitudinal and transverse frequencies of same tube.

Material.	Length cm.	D_m cm.	L/D_m .	N_L .	N_T .	N_L/N_T .	$0.4L/D_m$.
Brass . . .	70.74	2.420	29.23	2582	220.8	11.69	11.69
Copper . .	72.40	2.291	31.60	2672	210.5	12.70	12.64
Monel . . .	103.9	2.446	42.46	2226	132.6	16.79	16.98

There is sufficient agreement in the figures of the last two columns to warrant the conclusion that the ratio of the two frequencies can be expressed by equation (6), which also offers a means of calculating one of these frequencies if the other is known or can be measured.

The Partial Tones of Thin Cylindrical Tubes.

The mode of vibration of rods or bars vibrating laterally has been explained by Lord Rayleigh. For the fundamental of a "free-free bar" there are two nodes at a distance of 0.2242 of the length of the rod from

the ends, three nodes or two complete vibrating segments for the second partial, and so on. The distances of the nodes from the ends, for the 2nd, 3rd, 4th, 5th partials, etc., may be given by $0.661 L$ divided by 5, 7, 9, 11, etc., respectively, with the intermediate nodes equally spaced between these. The above was found to hold also for thin tubes. In a test of five thin brass tubes the nodal distances were calculated according to the above method, and marked and numbered. A tube was placed on rubber rests at the nodes, not necessarily the end ones, and the frequencies of the first five partials determined. In Table VIII. the frequencies of the fundamentals are given; the other partials are expressed as ratios of the fundamental. To ensure moderate accuracy over the range of frequencies involved, tubes were chosen with fundamental frequencies not too far from 200.

TABLE VIII.

Frequencies of first five partials of thin brass tubes.

Length cm.	N.	Frequency ratio.				
		N ₁ .	N ₂ .	N ₃ .	N ₄ .	N ₅ .
61.0	147	1	2.70	5.16	8.32	12.24
62.02	246	1	2.72	5.19	8.36	12.10
70.73	221	1	2.71	5.17	8.31	11.98
71.05	222	1	2.66	5.12	8.33	12.24
82.16	147	1	2.70	5.22	8.44	12.43
Mean	—	1	2.70	5.17	8.35	12.19

The theoretical frequency ratios of the partials of a "free-free" bar are 1, 2.76, 5.40, 8.93, 13.35, etc., forming a series which diverges at a greater rate than that found for tubes. Fenkner in 1879⁽⁷⁾ measured the frequencies of the peripheral vibrations of short cylinders of thin iron sheet open at one end and fixed at the other and set in vibration by short glass-rod generators. The frequencies were determined by means of Appunn reed tonometers, and his figures are for a "normal" cylinder of diam. 10.31 cm. and thickness about 0.037 cm. The frequency was independent of the height of the cylinder, and that of the fundamental had a value of about 96.

Fenkner's numbers were : 1, 2.77, 5.17, 8.33, 12.21.

In a later paper, Fenkner gave the series : 1, $\frac{8}{3}$, 5, 8, 12, 16, and Rayleigh, commenting on it, remarked that the agreement with the theoretical equation would be improved if these numbers were raised by about $1 \frac{1}{12}$ th part. The coincidence between Fenkner's figures for the series for his "normal" tube and those for the partials of brass

tubes is unmistakable, and indicates a close connection between the two modes of vibration.

Discussion and Results.

It is difficult to obtain tubes which are uniform in density, thickness and diameter, and the tubes employed were not perfect in these respects. This is shown, for example, by the variation in the weight per cm. of the tubes.

$$W/L = \text{Volume} \times \text{density} / \text{length} = \pi(R+r)(R-r)\rho,$$

and $(R+r)$ is the mean diameter, D_m , and $(R-r)$ the thickness, t , so that

$$W/L = \pi \rho t D.$$

The asymmetry of the tube of Table III. was noted in measuring the frequency of its peripheral vibrations; with one end free, the average frequency was 1890; with the other end free, the tone was duller and the frequency was 1900. Tubes may also suffer from an azimuthal asymmetry. In spite of these drawbacks, it has been possible to show that:

1. For a thin cylinder which is gradually shortened the product NL^2 is not constant, but diminishes slightly as the tube is shortened.

2. With tubes of equal length but different diameters the frequency is proportional to the diameter; for thin brass tubes 50 cm. long the frequency is approximately 180 times the mean diameter. For other materials the constant has to be multiplied by the ratio of the velocities of sound in the two materials.

3. The velocity of sound in the material of a cylindrical tube may be calculated from $V = \frac{NL^2}{1.259D_m}$, where N is the fundamental frequency of the transverse vibrations of the tube.

4. In brass tubes this law fails when the ratio of the length of the tube to its diameter is about 9, and is superseded by the law $N = \frac{kt}{D_m^2}$, where t is the thickness of the tube and k a constant of the material.

5. The frequencies of the fundamental longitudinal and transverse tones are in the ratio: $N_L/N_T = 0.4 \times \text{length}/\text{mean diameter}$.

6. The frequency ratios of the first five partials of thin cylindrical tubes are given by the series: 1, 2.70, 5.17, 8.35, 12.19. This coincides with the series found by Fenkner for the partials of short iron tubes sounding their peripheral tones.

Acoustical Applications of Metal Tubes.

The metal tube finds acoustical use as a dinner gong, as a unit in the chimes of an eight-day clock, and in the orchestral instrument called the glockenspiel, carillon or orchestral bells. It is also employed in a larger size as a church bell. In these instances use is not made of

the fundamental tone of the tube. A thick solid disk or short cylinder is fitted firmly into the upper end of the tube, and the tube is suspended from points or small holes at a distance of rather under a tenth of the length of the tube from the upper end and is struck at that end. The rich tone heard is the modified second partial of the tube accompanied by the fundamental which is feeble and soon dies out. The latter tone is an eleventh below the second partial and is not unpleasant. The addition of the inserted mass is equivalent to a lengthening of the tube at the loaded end with a consequent displacement of the nodes at that end, and the position of the node for the second partial becomes indefinite. In one respect the tube resembles a bell, and it emits the second partial although the tube is suspended from what, in the unloaded tube, would be the node for the third partial.

The measurements given in this paper were made, from time to time, over a period of years, in the Applied Physics Department of Glasgow University, and I have again to thank Dr. E. Geyer for his excellent reproduction of the graph.

Literature.

- (1) Rayleigh, vol. i. ch. viii.
- (2) Donkin, 'Acoustics,' ch. ix.
- (3) Violle, 'Cours de Physique,' vol. ii. pt. i.
- (4) C. Chree, Quart. Journ. Math. vol. xxiii., 1889.
- (5) A. Kalähne, *Ann. d. Phys.* (3) vol. xlv. p. 321, 1914.
- (6) G. E. Allan, *Phil. Mag.* (7) vol. xxvi., 1938.
- (7) H. Fenkner, *Ann. d. Phys.* (8) 1879.

LXVI. *Eigen-values and Eigen-Functions for the Operator $d^2/dx^2 - |x|$.*

By R. P. BELL, F.R.S. (Balliol College, Oxford)*.

[Received May 15, 1944.]

It is clear from general considerations that the operator $d^2/dx^2 - |x|^q$ will give rise to a discrete set of eigen-values and eigen-functions for any positive value of q . When $q=2$ the solutions involve the Hermitean polynomials, and receive an important application in the quantum theory of the harmonic oscillator. The case $q=4$ has been recently considered in connection with physical applications, though it is not possible to obtain explicit expressions for the eigen-values or eigen-functions (Bell, 1944). On the other hand, the case $q=1$ does not appear to have been treated, though it involves known functions and has possible physical applications.

* Communicated by the Author.

Solution of Equations.

The equation to be solved is

$$\frac{d^2\psi}{dx^2} - |x|\psi = -\lambda\psi, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

subject to the condition that ψ is continuous and of integrable square. The general solution of this equation is known in terms of Bessel functions (Watson, p. 97, equation (11)), and is

$$\psi = (\lambda - |x|)^{1/2} [A J_{1/3} \{ \frac{2}{3}(\lambda - |x|)^{3/2} \} + B J_{-1/3} \{ \frac{2}{3}(\lambda - |x|)^{3/2} \}]. \quad (2)$$

Let the eigen-functions for positive x be ψ_0^+ , ψ_1^+ , ψ_2^+ , etc., and those for negative x ψ_0^- , ψ_1^- , ψ_2^- , etc. Then it is easily seen that the appropriate form of the solutions is

$$\psi_n^+ = (-1)^n \psi_n^- = N_n (\lambda_n - |x|)^{1/2} [J_{1/3} \{ \frac{2}{3}(\lambda_n - |x|)^{3/2} \} + J_{-1/3} \{ \frac{2}{3}(\lambda_n - |x|)^{3/2} \}], \quad . \quad . \quad . \quad (3)$$

where N_n is a normalising factor. By considerations of symmetry ψ_n must be symmetrical about $x=0$ when n is even, and anti-symmetrical when n is odd. Moreover, when $|x| \rightarrow \infty$, (3) tends to zero, as may be seen by writing it in terms of the modified Bessel functions I and K (Watson, pp. 77-78),

$$\begin{aligned} \psi_n^+ &= (-1)^n \psi_n^- = N_n (|x| - \lambda_n)^{1/2} [I_{-1/3} \{ \frac{2}{3}(|x| - \lambda_n)^{3/2} \} - I_{1/3} \{ \frac{2}{3}(|x| - \lambda_n)^{3/2} \}] \\ &= \frac{N_n \sqrt{3}}{\pi} (|x| - \lambda_n)^{1/2} K_{1/3} \{ \frac{2}{3}(|x| - \lambda_n)^{3/2} \}, \quad . \quad . \quad . \quad (4) \end{aligned}$$

the asymptotic expression for $K_n(z)$ being $\lim_{z=\infty} K_n(z) = \sqrt{\frac{\pi}{2z}} e^{-z}$. It should be noted that the fractional indices in (3) and (4) do not introduce any real ambiguity, since when expanded the expressions for ψ contain only integral powers of x .

At $x=0$ the even eigen-functions must have $\frac{d\psi^+}{dx} = \frac{d\psi^-}{dx} = 0$, and the odd eigen-functions $\psi^+ = \psi^- = 0$. This gives immediately for the eigen-values λ_n ,

$$\left. \begin{aligned} J_{2/3}(\frac{2}{3}\lambda_n^{3/2}) - J_{-2/3}(\frac{2}{3}\lambda_n^{3/2}) &= 0 & n \text{ even,} \\ J_{1/3}(\frac{2}{3}\lambda_n^{3/2}) + J_{-1/3}(\frac{2}{3}\lambda_n^{3/2}) &= 0 & n \text{ odd,} \end{aligned} \right\} \quad . \quad . \quad . \quad (5)$$

using the relations (Watson, p. 45)

$$z \frac{d}{dz} \{ J_\nu(z) \} = \nu J_\nu(z) - z J_{\nu+1}(z) = -\nu J_\nu(z) + z J_{\nu+1}(z) \quad . \quad . \quad (6)$$

The solution can be completed by evaluating the normalising factor N_n from the condition $2 \int_0^\infty \psi_n^2 dx = 1$. Making the substitution $\frac{2}{3}(\lambda_n - |x|)^{3/2} = t_n$, this gives

$$12^{1/3} N_n^2 \int_0^{\frac{2}{3}\lambda_n^{3/2}} t_n^{1/3} \{ J_{1/3}(t_n) + J_{-1/3}(t_n) \}^2 dt_n = 1. \quad . \quad . \quad . \quad (7)$$

The integral need only be evaluated at $t_n = \frac{2}{3}\lambda_n^{3/2}$, since the integrand is real and positive throughout, and goes exponentially to zero at its other limit. To evaluate (7) we need the integrals

$$\left. \begin{aligned} \int^z z^{1/3} J_{1/3}^2(z) dz &= \frac{3}{2} z^{4/3} \{J_{1/3}^2(z) + J_{-2/3}^2(z)\}, \\ \int^z z^{1/3} J_{-1/3}^2(z) dz &= \frac{3}{2} z^{4/3} \{J_{-1/3}^2(z) + J_{2/3}^2(z)\}, \\ \int^z z^{1/3} J_{1/3}(z) J_{-1/3}(z) dz &= \frac{3}{2} z^{4/3} \{J_{1/3}(z) J_{-1/3}(z) - J_{2/3}(z) J_{-2/3}(z)\}, \end{aligned} \right\} \quad (8)$$

which are all special cases of the formulæ given by Lommel (1879). Equation (7) then becomes

$$2\lambda_n^2 N_n^2 [\{J_{1/3}(\frac{2}{3}\lambda_n^{3/2}) + J_{-1/3}(\frac{2}{3}\lambda_n^{3/2})\}^2 + \{J_{2/3}(\frac{2}{3}\lambda_n^{3/2}) - J_{-2/3}(\frac{2}{3}\lambda_n^{3/2})\}^2] = 1. \quad (9)$$

Taking into account (5) this gives for the normalising factor

$$\left. \begin{aligned} N_n &= [2^{1/2} \lambda_n \{J_{1/3}(\frac{2}{3}\lambda_n^{3/2}) + J_{-1/3}(\frac{2}{3}\lambda_n^{3/2})\}]^{-1} & n \text{ even,} \\ N_n &= [2^{1/2} \lambda_n \{J_{2/3}(\frac{2}{3}\lambda_n^{3/2}) - J_{-2/3}(\frac{2}{3}\lambda_n^{3/2})\}]^{-1} & n \text{ odd.} \end{aligned} \right\} \quad (10)$$

Numerical Values.

The odd eigen-values can be obtained directly from the zeroes of $J_{1/3}(z) + J_{-1/3}(z)$ tabulated by Watson (*loc. cit.*, p. 751)*. λ_2, λ_4 , etc., can be calculated with sufficient accuracy from the asymptotic formula given by Watson (*loc. cit.*, p. 506) for the zeroes of $J_\nu(z) \cos \alpha - Y_\nu(z) \sin \alpha$, which becomes $\frac{1}{2} \sec \frac{1}{2} \nu \pi \{J_\nu(z) - J_{-\nu}(z)\}$ on putting $\alpha = \frac{1}{2} \pi (\nu + 1)$. This leaves only λ_0 , which is best obtained by interpolation from tables of $J_{2/3}(z)$ and $J_{-2/3}(z)$. Existing tables (Dinnik, 1911) have an interval of 0.2 in z , and I have therefore calculated values at intervals of 0.02 from $z=0$ to $z=1$. In this range the ordinary Bessel function expansions give

$$J_{2/3}(z) = 0.69783 z^{2/3} (1 - 0.1500 z^2 + 0.00703 z^4 - 0.00016 z^6 + \dots) \quad (11)$$

$$J_{-2/3}(z) = 0.59255 z^{-2/3} (1 - 0.7500 z^2 + 0.07031 z^4 - 0.00251 z^6 + 0.000047 z^8 - \dots). \quad (12)$$

The values obtained are given in Table I., and interpolation gives $z=0.6856$ for the first zero of $J_{2/3}(z) - J_{-2/3}(z)$.

The first six eigen-values are given in the second row of Table II.

An asymptotic expression for the higher eigen-values is obtained by inserting in (5) the asymptotic expansions of the Bessel functions, when we find

$$\left. \begin{aligned} \cos \left\{ \frac{2}{3} \lambda_n^{3/2} + \frac{\pi}{12} \right\} &= \cos \left\{ \frac{2}{3} \lambda_n^{3/2} - \frac{7\pi}{12} \right\} & n \text{ even,} \\ \cos \left\{ \frac{2}{3} \lambda_n^{3/2} - \frac{5\pi}{12} \right\} &= -\cos \left\{ \frac{2}{3} \lambda_n^{3/2} - \frac{\pi}{12} \right\} & n \text{ odd,} \end{aligned} \right\} \quad (13)$$

* The second zero should be 5.5102, not 5.6102, as given in the Table.

in agreement with (14). The asymptotic values are given in the third row of Table II., and it is seen that with the exception of λ_0 they differ by less than 1 per cent. from the exact values.

It is of interest to compare the eigen-values of $d^2/dx^2 - |x^q|$ for different values of q , and the available data are shown in Table III. The case $q=4$ has been previously treated by a numerical method (Bell, 1944), and the values for $n > 1$ are less accurate than the remaining data. In the case $q = \infty$ we have to find values of λ such that the solution of $\left(\frac{d^2}{dx^2} + \lambda\right)\psi = 0$ is zero at $x = \pm 1$: these are, of course, $\lambda_n = \frac{1}{4}(n+1)^2\pi^2$.

TABLE III.

Eigen-values of $\left(\frac{d^2}{dx^2} - |x^q| + \lambda\right)\psi = 0$.

q .	λ_0 .	λ_1 .	λ_2 .	λ_3 .	λ_4 .	λ_5 .
1	1.019	2.338	3.247	4.088	4.819	5.521
2	1	3	5	7	9	11
4	1.061	3.800	7.36	11.40	16.27	21.22
∞	2.467	9.87	22.21	39.48	61.69	88.83

Physical Applications.

Equations (3) and (5) define the wave-functions and energy levels for a particle in a symmetrical potential energy well with $V(y) = a|y|$, a being a positive constant. The wave equation is

$$\frac{d^2\psi_n}{dy^2} + \frac{8\pi^2m}{h^2}\{W_n - a|y|\}\psi_n = 0, \quad (16)$$

which reduces to (1) on making the substitutions

$$x = \left\{\frac{8\pi^2ma}{h^2}\right\}^{1/3} y, \quad \lambda_n = W_n \left\{\frac{8\pi^2m}{h^2a^2}\right\}^{1/3} \quad (17)$$

Fig. 1 shows the first four wave functions (full curves) as defined by (5), (6) and (10): they are qualitatively similar to the harmonic oscillator functions. The difference is illustrated by the broken curve, which represents the lowest state of a harmonic oscillator having the same energy as the lowest state of $V(y) = a|y|$. A more likely physical situation is $V(y) = ay$ for $y > 0$, $V(y) = \infty$ for $y < 0$, corresponding to a charged particle in a uniform electrostatic field bounded by a sudden rise in potential energy. This problem has exactly the same solutions as $V(y) = a|y|$, except that only odd values of n are admissible.

However, a more useful application of Table III. is probably for interpolating the sequence of energy values for a potential well $V(y) = a_q|y^q|$, where q is between 1 and 4, but is not necessarily an integer. A wide variety of actual potential wells could be expressed empirically in

this form. The lower energy levels can be interpolated approximately from Table III., while the higher values are given with sufficient accuracy by an asymptotic expression. By making the substitutions

$$x=y\left\{\frac{8\pi^2ma_q}{\hbar^2}\right\}^{1/(q+2)}, \quad \lambda_n=\frac{W_n}{a_q^{2/(q+2)}}\left\{\frac{8\pi^2m}{\hbar^2}\right\}^{q/(q+2)}, \quad \dots \quad (18)$$

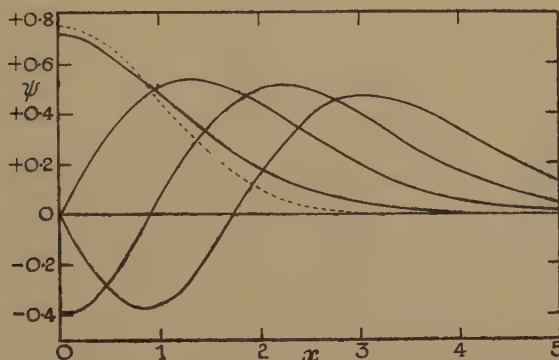
the Schrödinger equation becomes

$$\frac{d^2\psi_n}{dx^2} - |x^q|\psi_n = -\lambda_n\psi_n, \quad \dots \quad (19)$$

and a generalisation of (15) gives asymptotically

$$(2n+1)\pi = \oint (\lambda_n - |x^q|)^{1/2} dx, \quad \dots \quad (20)$$

Fig. 1.



and hence for large values of n

$$\lambda_n^{1/q+1/2} = \frac{q(n+1/2)\sqrt{\pi}\Gamma(1/q+3/2)}{\Gamma(1/q)} \dots \quad (21)$$

It is of interest to verify that equation (21) agrees with the Correspondence Principle relating the predictions of classical and quantum theory. It is easily shown classically that the mechanical frequency ν of an oscillator with $V(y)=a_q|y^q|$ and energy W_n is given by

$$\nu = \frac{q}{2\sqrt{2m\pi}} W_n^{1/2-1/q} a_q^{1/q} \frac{\Gamma(1/q+1/2)}{\Gamma(1/q)} \dots \quad (22)$$

We also have from (18) and (21)

$$\frac{1}{\hbar} \frac{dW_n}{dn} = \frac{q}{2\sqrt{2m\pi}} W_n^{1/2-1/q} a_q^{1/q} \frac{\Gamma(1/q+3/2)}{(1/q+1/2)\Gamma(1/q)} = \nu, \quad \dots \quad (23)$$

i. e., for large quantum numbers the frequency of the radiation emitted or absorbed in the transition $n \rightarrow n \pm 1$ becomes equal to the mechanical frequency of the oscillator.

References.

- Bell, *Proc. Roy. Soc. A* (1944). (October issue).
 Dinnik, *Arch. d. Math. u. Phys.* xviii. p. 337 (1911).
 Lommel, *Math. Ann.* xiv. p. 530 (1879).
 Dunham, *Phys. Rev.* xli. p. 713 (1932).
 Watson, 'A Treatise on the Theory of Bessel Functions' (Cambridge, 1922).

Balliol College,
 Oxford.

LXVII. *The Development of an Automatic Electric Balance for Research on Aerodynamic Stability.*

By J. B. BRATT, B.A., B.Sc., and K. C. WIGHT
 of the Aerodynamics Division, N.P.L.*

[Received April 26, 1944.]

1. *Introduction and Summary.*

The measurement of periodically varying forces is of considerable importance in experimental aerodynamics, especially in connection with problems on stability. The most general type of experiment envisaged is one in which a flexible model is inexorably forced at a number of points to produce a predetermined mode of motion, and the instantaneous reaction at each forcing point or earthed point is measured. Hitherto only simple cases of this problem have been attempted, in which a flexible aerofoil was forced at a single point.

Methods at present available for the measurement of the reactions are (a) optical, and (b) electrical. In the former method the reaction is applied to a spring, and the deflection recorded by means of an optical lever. A limitation of this method is that, for reasonable accuracy, the spring stiffness must be fairly low, and unless the mass of the system can be made very small, this results in a low value for the natural frequency, and hence a restricted frequency range over which accurate measurements can be obtained. Also large reactions may change the mode of motion considerably. The electrical method involves the use of some type of strain gauge or stress indicator, which may be of the resistance, capacity, or inductance type, and can be designed to have high stiffness. Since the electrical changes to be measured are very small, however, high gain amplifiers are required, and these must be of the D.C. type in order to avoid phase shift between electrical output

* Communicated by E. F. Relf.

and reaction and to enable static calibrations to be made. The disadvantages in this case lie in the complexity of the thermionic equipment, zero drifts in the D.C. amplifiers, and the necessity for frequent calibration on account of changes in the valve characteristics or operating conditions. There is the added difficulty of obtaining overall linearity of the measuring equipment, which is important if electrical analysis of the output is to be employed, whilst a further objection is the possibility of hysteresis in the sensitive unit. In the writers' experience this is certainly present in the resistance and inductance (magnetostriction) types.

In the experiment envisaged above it is easily possible to arrange for the measuring device to be stationary, and in this case its mass and size are of secondary importance. It is the purpose of this paper to describe an automatic electric balance which, when the above condition holds, seems ideally suited to this type of measurement. The balance is in use in the U.S.A. for the measurement of static and slowly varying forces having a period of about one second, and there is good reason to believe that it can be developed for measurement at 20 c.p.s. or more. Visual observation of the output current could be made with a cathode ray oscillograph as in the case of the electrical type of strain gauge.

As far as the authors are aware, no published account of the theory of action of the balance for periodic forces exists, and in view of this a theoretical treatment is included in this paper in order to bring out the main principles of design. It is shown that by suitable design the balance may be given the following properties:—

- (i.) high stiffness with corresponding good frequency response,
- (ii.) negligible phase shift,
- (iii.) linear calibration,
- (iv.) constant calibration.

The paper also includes an account of some tests made with a small experimental balance constructed at the N.P.L. The design of a full scale balance is now in hand, and it is proposed ultimately to use a number of these balances for the measurement of the complicated sets of forces such as arise during the flutter of flying wings or wings carrying several localized masses.

2. Description of Balance.

An automatic electric balance was developed in the U.S.A. by F. S. Eastman⁽¹⁾, and a six-component balance of this type is used in the large wind tunnel at the University of Washington for the measurement of static and slowly varying forces. The action of the balance is shown diagrammatically in fig. 1. A coil of wire C is suspended in the radial field of a permanent magnet M by means of parallel motion strips of negligible elastic stiffness arranged in such a manner that the coil is free to move in a direction normal to the magnetic flux. Displacement

of the coil alters the amount of light falling on a photocell which, by means of a simple, single-stage D.C. amplifier, controls the current flowing through the coil. When a load is applied to the coil the latter is displaced until the reaction due to the current balances the total load. There is thus an exact proportionality between current and load provided that the reaction due to the suspension is negligible and that the magnetic field acting on the coil is unchanged for the displaced position. These conditions are satisfied if the displacement of the coil is very small, which is equivalent to the "electrical" stiffness of the system being high. Constancy of calibration is attained by the use of a permanent magnet. The current-load proportionality is, within wide limits, clearly independent of non-linearity in the amplifier and photocell

Fig. 1.

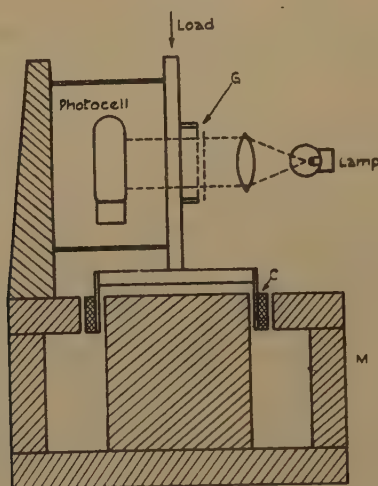


Diagram of balance.

control system or changes in sensitivity or gain due to ageing of characteristics and variations in applied voltages.

The light control used in the American balance consists of a small mirror which is tilted by displacement of the coil and deflects a beam of light falling on the photocell. This method was tried initially in the small experimental balance at the N.P.L., but was finally abandoned in favour of the pair of gratings *G* indicated in fig. 1. These gratings were produced photographically and have 200 lines to the inch, the width of the lines being equal to the width of the clear spaces between them. One of the gratings is fixed, the other being attached to the spindle carrying the coil, and a displacement of the latter by 0.0025 in. gives the maximum change in the amount of light transmitted. This method of light control by very small displacements has been used by Cristescu⁽²⁾ and others, and works extremely well. When used with a

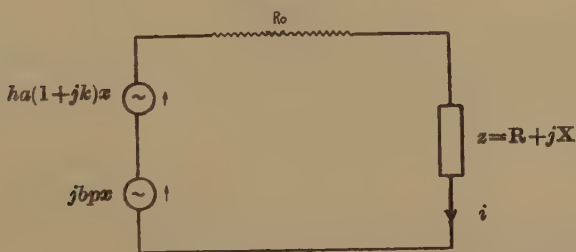
CM vacuum type photocell and a 48-watt lamp it is possible to obtain from 4 to 8 microamps. from the cell per 0.001 in. displacement.

The N.P.L. experimental balance was built around a loud speaker magnet, and since the gap of the latter was very small, the electrical stiffness and load capacity were low. The elastic stiffness of the suspension was approximately 10 per cent. of the total stiffness. The capacity of the balance was ± 40 gm., and the output current for this loading was about ± 30 milliamps. with a mean current of 40 milliamps. Corresponding displacements of the coil were approximately ± 0.0005 in.

3. Theory for Sinusoidal Load.

It is assumed in the following analysis that the applied load P , displacement of the coil x , and output current i all vary sinusoidally with time and have an angular frequency p , whilst all circuits are regarded as being linear. The equivalent circuit for the output valve of the amplifier is shown in fig. 2 in which complex notation is employed*.

Fig. 2.



Equivalent circuit of amplifier output valve.

Two generators are shown inserted in this circuit, one corresponding to the input at the grid of the valve and the other to the back E.M.F. in the coil. It will be shown later that the former E.M.F. is of the form $ha(1+jk)x$, where h and k are functions of frequency such that h tends to unity and k tends to zero when p tends to zero. If ϵ is the phase shift between displacement and input, then $k = \tan \epsilon$. The back E.M.F. is proportional to the velocity of the coil and is in the same sense as the input, since both tend to produce current opposing the motion of the coil. The output valve impedance is R_0 , and the additional impedance due to the coil and measuring instruments, etc., is represented by $Z = R + jX$, where X is a reactance which tends to zero with p .

It follows from fig. 2 that

$$\begin{aligned} i &= \frac{ha(1+jk) + jbp}{R_0 + R + jX} x \\ &= \frac{ha}{R_0 + R} \beta x, \dots \dots \dots (1) \end{aligned}$$

* Throughout this paper $\sqrt{-1}$ is denoted by j in accordance with alternating current theory.

where
$$\beta = \frac{1+j\left(k+\frac{pb}{ha}\right)}{1+\frac{jX}{R_0+R}} \quad \dots \dots \dots (2)$$

The expression β accounts for phase shift in the relation between x and i , and tends to unity as p tends to zero. The numerator introduces the effect of back E.M.F. in the coil and electrical phase shift preceding the output stage of the amplifier, whilst the denominator accounts for the effect of reactance in the output circuit.

If γ is the coil reaction per unit current, N the number of turns in the coil, r its mean radius, and H the field strength of the magnet,

$$\gamma = 2\pi rNH = b, \quad \dots \dots \dots (3)$$

where all quantities are expressed in C.G.S. units. It will be assumed that the elastic stiffness of the balance is negligible compared with the electrical stiffness, and if the value of the latter for static loading is σ_e , then

$$\sigma_e = \frac{\gamma i}{x}, \quad \dots \dots \dots (4)$$

where i and x have corresponding values for $p=0$, and since for this condition $h=\beta=1$, it follows from (1) that

$$a = \frac{\sigma_e(R_0+R)}{\gamma} \quad \dots \dots \dots (5)$$

Substitution from (3) and (5) in (1) and (2) gives

$$i = \frac{h\sigma_e}{\gamma} \beta x, \quad \dots \dots \dots (6)$$

and
$$\beta = \frac{1+j\left\{k+\frac{p\gamma^2}{h\sigma_e(R_0+R)}\right\}}{1+\frac{jX}{R_0+R}} = u + jv. \quad \dots \dots \dots (7)$$

If air damping is assumed to be negligible, the equation of motion of the coil is

$$mD^2x + \gamma i = P, \quad \dots \dots \dots (8)$$

where m is the mass of the coil and attached structure and D denotes the usual differential operator d/dt . The solution of (8) for the steady state is

$$-mp^2x + \gamma i = P,$$

or
$$P = \gamma i \left(1 - mp^2 \frac{x}{\gamma i}\right). \quad \dots \dots \dots (9)$$

The dynamic electrical stiffness σ_e' may be derived from (6) and is expressed by

$$\sigma_e' = h\sigma_e u, \quad \dots \dots \dots (10)$$

where \bar{u} is the real part of β . Hence the natural frequency of the balance is given by

$$p_0^2 = \frac{\sigma_{e0}'}{m} = \frac{h_0 \sigma_{e0} u_0}{m}, \quad \dots \dots \dots (11)$$

provided the damping forces due to phase difference between reaction and displacement are not large in relation to the stiffness forces*. The zero suffix indicates values of p , σ_e' , h and u corresponding to the natural frequency.

Substitution from (6) and (11) in (9) gives finally

$$P = \gamma i \left(1 - \frac{p^2}{p_0^2} \cdot \frac{h_0 u_0}{h \beta} \right) \dots \dots \dots (12)$$

It follows from this expression that phase shifts in the electrical circuits are not involved directly in the relation between P and i , but operate on the term $\frac{p^2}{p_0^2}$ which corresponds to the usual frequency response term of a vibratory system and in a properly designed balance will be very small. It is to this fact that the possibility of developing the balance for the measurement of periodic forces is due.

Since all the equations in the analysis are linear it follows that the response to a periodic load of general wave form may be obtained by application of the above formulæ to each of its harmonic components in turn. No attempt is made in this paper to estimate the effect of non-linearities in the control circuits, but it is clear from the nature of the balance that the effect can only arise through a frequency response term, which is likely to be small. This is supported by the results of tests made with the N.P.L. experimental balance, the valve circuits of which were certainly not linear. Where the output valve is a pentode or tetrode improved linearity without loss of output may be obtained by the inclusion of resistance in the anode circuit sufficient to make the total load approximately equal to the recommended value. This may also be necessary to avoid overloading the valve.

It is necessary to exercise considerable care in design if the amplifier is to be energized from the A.C. mains by means of a power unit, since the smoothing circuit of the latter is included in X , which may become very large if the circuit resonates to the frequency at which measurements are being made. The difficulty can be avoided, however, by the use of a somewhat more complicated amplifier of the push-pull type.

4. Stability of Balance.

The reaction opposing motion in a vibratory system is of the form

$$(A + jBp)x, \quad \dots \dots \dots (13)$$

where A is a stiffness and B is the damping. For positive damping, therefore, the reaction is in advance of the displacement. Since in the

* More exactly $p_0^2 = \frac{\sigma_{e0}'}{m} \left(1 - \frac{1}{4} \tan^2 \phi \right)$ where ϕ is the phase difference.

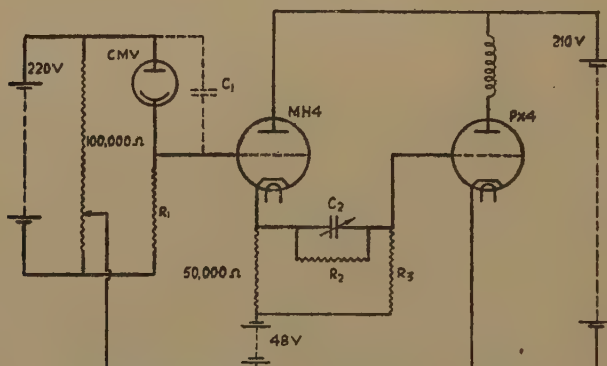
automatic balance the reaction is electrical, the possibility exists of its lagging behind the displacement at the natural frequency of the system. This corresponds to negative damping and the balance would be unstable.

It follows from (6) and (7) that the condition for stability is

$$k_0 + \frac{p_0 \gamma^2}{h_0 \sigma_e (R_0 + R)} > \frac{X_0}{R_0 + R}, \quad \dots \dots \dots (14)$$

where k_0 and X_0 correspond to the natural frequency of the system. The value of k_0 for the conventional photocell and amplifier circuit is negative and has been found to be numerically greater than the term following for the required practical values of the quantities involved unless the electrical stiffness is low, whilst X_0 is positive. The condition for stability is, therefore, not satisfied, a fact which was verified when the N.P.L. experimental balance was first tried out. It is undesirable to obtain stability by the application of some form of damping to the

Fig. 3.



Amplifier circuit for experimental balance.

$R_1 = 5\text{M}\Omega$; $R_2 = R_3 = 0.25\text{M}\Omega$; $C_1 = 90\mu\text{F}$; $C_2 = 0.001\mu\text{F}$.

coil, since this would introduce a direct phase shift into the relation between P and i . This may be seen from (8), which now includes a term μDx , and leads to the relation

$$P = \gamma i \left\{ 1 + \left(j \frac{\mu p}{\sigma_{e_0}} - \frac{p^2}{p_0^2} \right) \frac{h_0 u_0}{h \beta} \right\}. \quad \dots \dots \dots (15)$$

The method adopted for the balance mentioned above was to change the sign of k_0 and control its magnitude by means of a phase-shifting network introduced into the amplifier circuit. This will be described in the following section. Stability was attained in the American balance by the application of fluid damping.

When the N.P.L. balance was first set up the light control was obtained by means of a tilting mirror mounted on spring strip pivots. It was found, however, that the hysteresis in the spring strips introduced

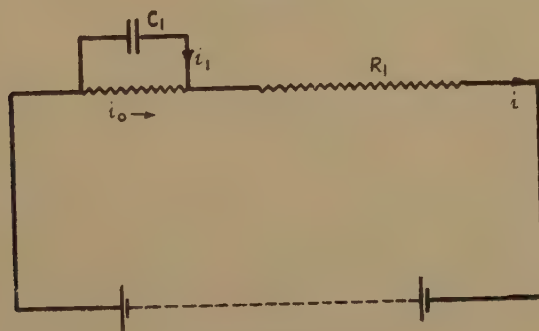
a phase shift in a direction tending to produce instability, and the mirror was replaced by the pair of gratings already described.

5. Photocell and Amplifier Circuits.

The circuit used in conjunction with the N.P.L. experimental balance is illustrated in fig. 3. Since the required photocell load resistance R was very much higher than the maximum permissible resistance for the grid circuit of the output valve, a buffer stage was introduced, which consisted of an MH4 used as a cathode follower. This valve was found to work satisfactorily with a high resistance in its grid circuit, and enabled the phase-shifting network $C_2R_2R_3$ to be introduced in a convenient manner.

The output from the photocell circuit may be derived from fig. 4, in which C_1 is the capacity of the cell and i_0 is the emission current. Provided the cell is of the vacuum type and is working at saturation,

Fig. 4.



Photocell circuit.

i_0 will be proportional to and in phase with the light variation and hence with the coil displacement. Since the voltage drop across the cell is equal and opposite to the drop across R_1 ,

$$\frac{i_1}{j\omega C_1} = -R_1 i, \quad \dots \dots \dots (16)$$

$$\text{also} \quad i = i_0 + i_1, \quad \dots \dots \dots (17)$$

hence by elimination of i_1 between (16) and (17)

$$i = \frac{i_0}{1 + j\omega C_1 R_1}, \quad \dots \dots \dots (18)$$

and the output is given by

$$R_1 i = \frac{R_1 i_0}{1 + j\omega C_1 R_1}. \quad \dots \dots \dots (19)$$

It follows from (19) that the output lags behind the displacement, and on account of the large value of R_1 this lag may be appreciable (16° at 100 c.p.s. for the values given in fig. 3).

The impedance of the phase shifting network shunting the cathode circuit resistance of the cathode follower is $R_3 + Z_2$, where

$$Z_2 = \frac{R_2}{1 + jpC_2R_2} \quad \dots \quad (20)$$

If this impedance is large compared with the cathode circuit resistance, and V is the voltage across the latter, the output from the network is

$$\frac{VR_2}{R_3 + Z_2} = V\alpha \frac{1 + jpC_2R_2}{1 + jpC_2R_2\alpha}, \quad \dots \quad (21)$$

where $\alpha = R_3/(R_2 + R_3)$. Since α is less than unity (21) introduces a positive phase shift. It follows from (19) and (21) that the E.M.F. in the equivalent circuit of the output valve corresponding to the input at the grid may be written

$$KR_1\alpha \frac{1 + jp\tau_2}{(1 + jp\tau_1)(1 + jp\tau_2\alpha)} \cdot x, \quad \dots \quad (22)$$

where $\tau_1 = C_1R_1$ and $\tau_2 = C_2R_2$. This is of the form $h\alpha(1 + jk)x$ introduced in §3, and substitution of $p=0$ gives $KR_1\alpha = a$, hence

$$h(1 + jk) \frac{1 + jp\tau_1}{1 + jp\tau_1(1 + jp\tau_2\alpha)} \quad \dots \quad (23)$$

Excellent control of the damping could be obtained in the experimental balance by the use of a variable condenser for C_2 , and changeover from instability to heavy positive damping was well within the range of a small commercial type condenser.

The phase shift introduced by the interelectrode capacities of the valves is in the same sense as that due to the photocell, but it may be shown to be comparatively very small, provided the Miller effect in the output valve is negligible. This was assured in the experimental balance by a low anode load and may be obtained in a larger balance by the use of a pentode or tetrode.

The phase shift obtainable from the network has a maximum value for a given value of α . If θ is the phase shift, then from (21)

$$\theta = \tan^{-1} p\tau_2 - \tan^{-1} p\tau_2\alpha,$$

$$\text{or} \quad \tan \theta = \frac{p\tau_2 - p\tau_2\alpha}{1 + p^2\tau_2^2\alpha} \quad \dots \quad (24)$$

This is a maximum when $p\tau_2 = 1/\sqrt{\alpha}$, and hence

$$(\tan \theta)_{\max} = \frac{1 - \alpha}{2\sqrt{\alpha}} \quad \dots \quad (25)$$

Thus the maximum shift attainable increases as α is reduced, but this is offset by a reduction in the gain of the amplifier.

6. Tests with the Experimental Balance.

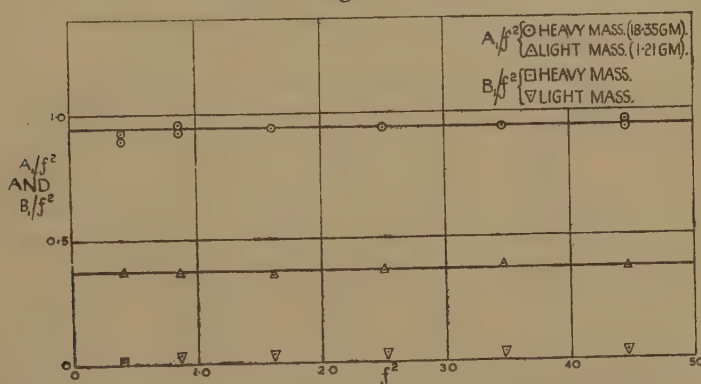
In these tests a horizontal arm was pivoted at one end to the spindle of the balance coil, and a point on the arm two-thirds of its length from

this end was given a sinusoidal motion in a vertical plane by attachment to a mechanical oscillator. This attachment was made by means of a link in order to avoid end constraint on the arm. The free end of the arm carried a mass.

With this arrangement an inertial reaction was applied to the balance in phase with the motion of the oscillator and proportional to the square of the frequency. The in-phase and out-of-phase components of this reaction, as indicated by the balance, were then measured by means of an electrical harmonic analyser. Two sets of tests were made, one with a heavy mass and the other with a light mass attached to the arm, and in each case the frequency f ranged from 2.00 to 6.67 c.p.s. The virtual inertia of the system was maintained constant by the use of masses of the same dimensions but different densities.

The results of these tests are shown in fig. 5, in which the values of A_1/f^2 and B_1/f^2 are plotted against f^2 , where A_1 and B_1 are the in-phase

Fig. 5.



Frequency response for the N.P.L. experimental balance.

and out-of-phase components respectively expressed in arbitrary units.

Theoretically the quantities plotted are proportional to $1 + \frac{f^2}{f_0^2}$, provided $\frac{f^2}{f_0^2}$ is small, where f_0 is the natural frequency of the system. Hence the points should lie on straight lines, and in the case of the heavy mass the rise in response at a frequency of 6.67 c.p.s. should be $\frac{6.67^2}{82^2}$ since the

natural frequency was 82 c.p.s. This is about $\frac{1}{2}$ per cent. and is barely distinguishable on the scale of fig. 5. The rise in response for the light mass would be still less. The results, therefore, appear to be in good agreement with theory as far as frequency response is concerned.

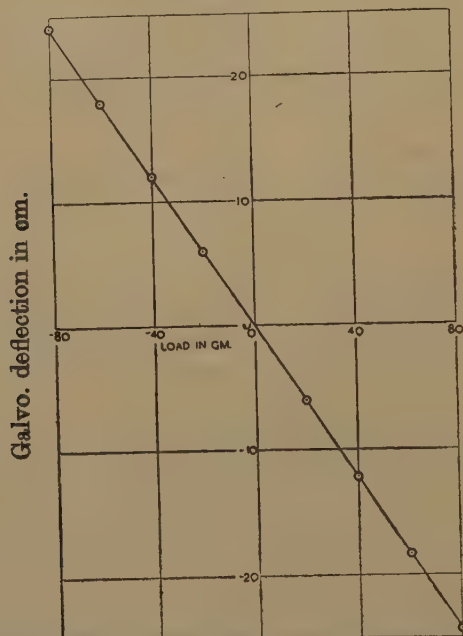
The values of B_1/f^2 for the two tests are very small and practically coincide. This coincidence suggests that the out-of-phase component cannot be due to phase shift in the balance to any appreciable extent, but results from still air damping and friction in the pivots. Shift in

the balance must certainly be less than 1° , otherwise the measured values of B_1/f^2 for the heavy mass would be greater.

A calibration curve for the balance is given in fig. 6, and by means of this the displacement between the two curves in fig. 5 has been determined in gramme-second units. The value of 0.415 obtained agrees extremely well with the value 0.413 calculated from the known difference of mass and amplitude of oscillation.

The results of these tests, together with the implications of theory, are regarded as a satisfactory indication of the possibility of the development of the balance for the measurement of periodic forces where phase

Fig. 6.



Calibration curve for N.P.L. experimental balance.

relationships are to be preserved. Calculations based on the treatment given in the previous sections, and taking into account the technical limitations in the construction of large permanent magnets, etc., show that a balance having the following characteristics is possible.

Load capacity	± 5 lb.
Static stiffness	10,000 lb./in.
Mass of coil and attachments	4.0 lb.
Natural frequency	200 c.p.s.
Phase shift between load and current at 20 c.p.s.	less than 0.1°
Output per lb. load	10 milliamps.
Frequency response at 20 c.p.s.	1% high.

In these calculations it has been assumed that the phase-shifting network is adjusted to give maximum shift at the natural frequency of the balance. Minimum shift between load and output current would be obtained by a reduction of the shift in the network to the point where the balance is just stable.

It is interesting to note that the stiffness quoted above is approximately equal to that of a steel wire 6 in. long and $\frac{1}{20}$ in. in diameter.

The proposed amplifier for this balance has push-pull output with the coil connected across the anodes of the output valves. This has the advantage of making the mean current through the coil zero, and if the amplifier is energized from the mains, avoids effects due to the inclusion of the power unit smoothing circuit in the output circuit (see § 3). The application of the foregoing theoretical treatment to this type of circuit results in the replacement of R_0 by $2R_0\delta$ in the derived formulæ, where $\delta=R_1/(R_1+R_0)$ and R_1 is the load resistance in the anode circuit of each output valve.

Prediction of stability depends partly upon a knowledge of the reactance in the output circuit, and this is difficult to estimate. In the present case, if the coil is assumed to form the major part of the reactance, this will be inductive and would have a value of approximately 1.3 henrys for an air core. The presence of the magnetic material will probably increase the inductance by a factor lying between 0 and 10 approximately, whilst eddy currents will tend to oppose this increase. The balance would be perfectly stable for this value, but in cases where coils of higher inductance are used it would be advisable to include in the circuit a coil exactly similar to the moving coil and fixed immediately beneath the latter on the core of the magnet. If the coils are connected so that their fluxes are in opposition, the inductance of the combination will be very small, and in addition the demagnetizing effect on the magnet will be reduced to a similar extent.

The formulæ required in connection with the design of a balance with the simple amplifier circuit of fig. 3 are collected below and are most conveniently applied if a natural frequency is first assigned to the balance.

$$\gamma=2\pi rNH,$$

$$\sigma_e=\frac{\gamma^2}{x},$$

$$\sigma'_e=h\sigma_e u,$$

$$h(1+jk)=\frac{1+jp\tau_2}{(1+jp\tau_1)(1+jp\tau_2\alpha)},$$

$$\beta=-\frac{1+j\left\{k+\frac{p\gamma^2}{h\sigma_e R_0+R}\right\}}{1+\frac{jX}{R_0+R}}=u+jv,$$

$$P = \gamma i \left(1 - \frac{p^2}{p_0^2} \cdot \frac{h_0 u_0}{h \beta} \right),$$

$$k_0 + \frac{p_0 \gamma^2}{h_0 \sigma_e (R_0 + R)} > \frac{X_0}{R_0 + R} \text{ for stability.}$$

The work described above was carried out in the Aerodynamics Division of the National Physical Laboratory, and it is published on the recommendation of the Aeronautical Research Committee and by permission of the Director of the Laboratory.

References.

- (1) F. S. Eastmen, "The Electromagnetic Balance—A High-Precision Measurement and Control Device", 'Instruments', vol. xiv. p. 284 (Oct. 1941).
- (2) G. D. Cristescu, "Photoelektrische oder thermoelektrische Verfahren zur Feststellung und Messung sehr kleinen Verschiebungen", Physik. Zeitschr. vol. xxx. p. 24 (1929).

LXVIII. *Influence of Air-gases on the First and Second Positive Systems of Nitrogen.*

By N. R. TAWDE, M.Sc. (Bom.), Ph.D. (Lond.), F.Inst.P., F.N.I., and
V. S. PATANKAR, M.Sc., Ph.D. (Bom.), Royal Institute of Science,
Bombay *.

[Received February 7, 1944.]

Introduction.

THE addition of a foreign gas to another, whose spectrum is to be examined, has very often the most remarkable effects upon the lines of bands developed. Among these the noble gases, as diluents, generally alter the intensity distribution to a great extent. In some cases they are known even to bring out spectra which are not otherwise obtainable. It is known that atmospheric dry air contains, besides the principal constituent nitrogen, variable amounts of oxygen, carbon dioxide, hydrogen and traces of argon, neon, helium, krypton and xenon, and when excited in a discharge-tube it gives the characteristic and well-known N_2 band-systems. It is expected, therefore, that these band-systems, because of the admixture of several rare gases, will show a different appearance from that given by pure undiluted nitrogen. With this idea in view we have carried out a comparative study of vibrational intensity distribution among the First and Second Positive band-systems of N_2 as given by pure nitrogen gas and by air.

* Communicated by the Authors.

Experimental.

The spectra were excited in a discharge-tube of the usual H-type fitted with aluminium electrodes and were examined end-on through the quartz window of the tube. The tube was 25 cm. long and 7 mm. in diameter, and had two small bulbs containing P_2O_5 and KOH to remove water vapour and carbon dioxide. The pressure of the enclosed gas was adjusted to 1.2 mm. The tube was run by an induction coil

TABLE I.

N_2 First Positive System; Peak Intensities with Nitrogen (I_N) and Air (I_A).

Band.	I_N .	I_A .	I_N/I_A .
5, 2 ($\lambda=6703$)	24.63	10.71	0.44
6, 3	34.34	15.30	0.45
7, 4	20.61	9.01	0.44
8, 5	13.25	5.96	0.45
9, 6	8.73	4.25	0.49
10, 7	4.77	2.39	0.50
11, 8	2.31	1.19	0.52
12, 9	1.73	0.80	0.46
5, 1 ($\lambda=6126$)	3.75	1.45	0.39
6, 2	6.12	2.67	0.44
7, 3	6.35	2.72	0.43
8, 4	6.01	2.56	0.43
9, 5	5.70	2.49	0.44
10, 6	4.92	2.38	0.48
11, 7	5.03	2.31	0.46
12, 8	3.80	1.82	0.48
6, 1 ($\lambda=5593$)	0.68	0.14	0.20
7, 2	0.98	0.26	0.26
8, 3	1.10	0.38	0.35
9, 4	1.49	0.55	0.37
10, 5	2.36	0.85	0.36
11, 6	2.79	1.15	0.41
12, 7	2.86	1.11	0.39

giving a high-frequency voltage, no condenser or spark gas being included in the circuit; the primary voltage was 10 v. and with it the current in the secondary was 1.0 ma.

The technique of intensity measurements followed by us is the same as that used and described in detail by Read and Johnson⁽¹⁾, Johnson and Tawde⁽²⁾, Elliott and Cameron⁽³⁾ and others. The details of other experimental procedure can also be found in our recent paper⁽⁴⁾.

Results.

It is known that the intensities when evaluated by integration methods give a true measure of the relative band-intensities in a system. For shortening the experimental procedure, peak intensity values are generally

taken as proportional to the integrated intensity values⁽⁵⁾. We have, however, shown recently⁽⁶⁾ that, at least in the N₂ Second Positive

TABLE II.
N₂ Second Positive System ; Peak Intensities with
Nitrogen (I_N) and Air (I_A).

Band.	I _N .	I _A .	I _A /I _N .
2, 8	2.23	0.75	0.34
3, 9	2.03	0.94	0.46
0, 5 ($\lambda=4665$)	2.90	1.36	0.47
4, 10	1.67	0.81	0.48
1, 6	4.66	3.69	0.80
2, 7	4.18	3.45	0.83
3, 8	2.89	2.21	0.77
0, 4 ($\lambda=4344$)	10.04	4.29	0.43
1, 5	21.69	8.63	0.40
2, 6	19.18	9.80	0.51
3, 7	11.33	5.20	0.46
4, 8	6.45	3.14	0.49
0, 3 ($\lambda=4059$)	57.62	22.78	0.40
1, 4	73.94	35.84	0.49
2, 5	46.31	19.54	0.42
3, 6	22.91	12.98	0.57
4, 7	8.35	4.34	0.52
0, 2 ($\lambda=3805$)	189.0	115.8	0.61
1, 3	162.4	108.1	0.67
2, 4	71.88	44.20	0.61
3, 5	22.46	15.22	0.68
4, 6	4.51	2.79	0.62
0, 1 ($\lambda=3577$)	590.2	489.8	0.86
1, 2	183.6	152.3	0.83
2, 3	46.28	25.40	0.55
0, 0 ($\lambda=3371$)	1000	1000	1.00
1, 1	42.60	23.44	0.55
2, 2	29.68	15.78	0.53
3, 3	49.16	35.67	0.73
4, 4	32.21	17.82	0.55
1, 0 ($\lambda=3159$)	883.7	922.1	1.04
2, 1	430.8	282.7	0.66
3, 2	153.2	96.49	0.63
4, 3	56.50	45.02	0.80
2, 0 ($\lambda=2977$)	350.5	220.9	0.63
3, 1	337.5	226.4	0.67
4, 2	220.8	141.1	0.64

system, this proportionality does not exist. The present investigation aims at establishing the difference, if any, in the distribution of intensities in a system or systems of the same molecule when excited alone and in the presence of another gas, other conditions being exactly identical. Hence it is permissible to use the peak intensity values, which are much

simpler to estimate experimentally, and it is not necessary to follow the very laborious integration method.

Accordingly the peak intensity values of the N_2 First and Second Positive systems are determined and given in Tables I. and II., the values for the individual bands in both the systems being expressed on a relative scale, 1000 units for the (0,0) band of the Second Positive system. Columns 2 and 3 of the tables give the values of I_N and I_A , the peak intensities in pure nitrogen and in air, respectively; and column 4 gives the calculated values of the ratio of I_A/I_N .

Discussion and Conclusions.

In Tables I. and II. it is seen that the ratio I_A/I_N , though somewhat steady for some bands, shows fluctuations between 0.20 and 1.04; this observation implies a large range of variation, the limits of experimental errors in determining the band intensities being less than 10 per cent. This means that the relative intensity distribution of bands tends to alter appreciably as discharge conditions are changed from pure nitrogen to air-gases. If the ratio I_A/I_N had remained constant within the limits of experimental errors, it would have been justifiable to assume the ineffectiveness of small amounts of air-gases to have no effect on the spectra of the chief constituent N_2 ; the range of variation is too large, however, to justify any such assumption. The different aspects of this variation will therefore be studied below.

I. First Positive System.

The values in column 4 of Table I. show the change of band-intensities obtained from the discharge in air and in nitrogen; the average ratio over the whole range of observed bands is 0.42. The selective enhancement or weakening of certain progressions can now be studied by reference to Table III. and the graph (fig. 1). It is seen that there is a selective

TABLE III.

First Positive System.

v'' -progression with $v' =$	5	6	7	8	9	10	11	12
Average value of I_A/I_N	0.27	0.36	0.38	0.41	0.43	0.45	0.46	0.44

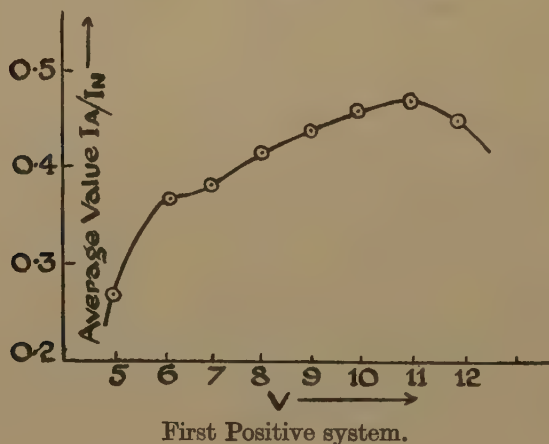
enhancement in the neighbourhood of $v'=11$ of the upper ($B^3\Pi$) state, and a less distinct but a definite enhancement at $v'=6$; this is noticed only on a close examination of the values I_A/I_N from Table I. This means that the N_2 radiations obtained from discharge in air are subject to influence from the neighbouring constituents.

The result is significant in view of the fact that the First Positive bands given by active nitrogen exhibit a similar effect^{(7), (8)}. We may, therefore, postulate that active nitrogen is somehow produced to some extent, within the uncondensed discharge through air; the special

enhancement of bands near $v' \sim 11$ and $v' \sim 6$ in the $B^3\Pi$ state is then a consequence which should follow from this hypothesis. For an explanation as to how an uncondensed discharge in air gives results akin to those of active nitrogen (which is generally produced in a condensed discharge in pure nitrogen), it is necessary to examine the influence exerted by the neighbouring molecules in air foreign to its nitrogen constituent. As foreign matter is essential for the production of active nitrogen and as inert gases are chiefly responsible for bringing about a redistribution of spectral intensities, the effect may be explained on the hypothesis that it is perhaps argon, the major rare gas constituent of air, that plays some part in activating the chief nitrogen component.

The exact rôle played by foreign gases is still a matter in dispute, but they may function in such a way as to accelerate the formation of excited molecules from atoms, or to retard the rapid reversion of active

Fig. 1.



nitrogen, as demanded by one or other of the theories of active nitrogen due to Sponer and Herzberg⁽⁹⁾, Okubo and Hamada⁽⁸⁾, and Cario and Kaplan⁽¹⁰⁾.

These theories predict enhancements at $v' \sim 6$ and $v' \sim 11$, the effect at $v' \sim 6$ being expected to be the greater of the two. Our experimental results, on the other hand, reveal a very distinct enhancement at $v' \sim 11$ and a slight trace of one at $v' \sim 6$. Hence these effects noticed by us and their agreement with previous experimental results obtained by others in active nitrogen, provide yet another proof of the unconvincing nature of the active nitrogen theories, which have been known to satisfy the experimental results only partially.

II. Second Positive System.

The selective effect produced by air gases is also noticed in the case of the Second Positive system. Table IV. shows the average values of the ratio I_A/I_N for the different v'' -progressions of this system.

This table shows fluctuations in the ratio I_A/I_N with selective enhancement in the neighbourhood of $v'=0$ and 1. An examination of the values I_A/I_N in Table II. reveals that in air the bands (1, 0), (0, 0), (1, 2), (0, 1), (3, 5), (1, 3), (0, 2), (3, 8), (2, 7) and (1, 6) are relatively enhanced as compared with other bands. An examination of the greater part of the afterglow spectrum of active nitrogen, photographed by Fowler and Strutt⁽¹¹⁾, shows that in the Second Positive system only a limited number of bands appear, namely, the (1, 0), (0, 0), (1, 2), (0, 1), (1, 3) and (0, 2) bands. It will thus be noticed that the bands which show enhancement in the discharge through air in our experiments are those which selectively appear in the afterglow spectrum of nitrogen and a few more in addition. These effects lead us to the same conclusion as before, *i. e.*, traces of active nitrogen are produced by the passage of an uncondensed discharge through air.

TABLE IV.

Second Positive System.

v' -progression with $v' = \dots\dots\dots$	0	1	2	3	4
Average value of $I_A/I_N \dots\dots\dots$	0.63	0.68	0.56	0.62	0.59

In another paper it is proposed to study and compare the intensity distribution in the First Positive system with that in the Second Positive, when the discharge is run through air and pure nitrogen.

The authors take this opportunity to express their sincere thanks to Sir C. V. Raman, Kt., F.R.S., N.L., for offering us the facilities of the use of the micro-photometer at the Indian Institute of Science, Bangalore.

References.

- (1) Read and Johnson, *Phil. Mag.* xi. p. 1152 (1931).
- (2) Johnson and Tawde, *Proc. Roy. Soc. cxxxvii. A*, p. 575 (1932).
- (3) Elliott and Cameron, *Proc. Phys. Soc. xlv. p. 801* (1934).
- (4) Tawde and Patankar, *Proc. Phys. Soc. lv. p. 396* (1943).
- (5) Ornstein and Brinkman, *Proc. Amsterdam*, xxxiv. p. 1 (1931).
- (6) Tawde and Patankar, *Bombay University Journal*, vol. ix. pt. 3 (1940).
- (7) Jevons, *Report on Band Spectra of Diatomic Molecules*, p. 207.
- (8) Okubo and Hamada, *Phys. Rev.* xlii. p. 795 (1932).
- (9) H. Sponer, *Zeit. f. Phys.* xxxiv. p. 622 (1925).
G. Herzberg, *Zeit. f. Phys.* xlix. p. 512 (1928).
H. Sponer, *Zeit. f. Phys.* lxxxix. p. 431 (1934).
- (10) Cario and Kaplan, *Zeit. f. Phys.* lviii. p. 769 (1929).
Cario, *Zeit. f. Phys.* lxxxix. p. 523 (1934).
- (11) Fowler and Strutt, *Proc. Roy. Soc. lxxxv. A*, 377 p. (1911).

15th September, 1943.

LXIX. *An Electrical Circuit for Harmonic Analysis and other Calculations.*

By T. B. RYMER, Ph.D., and C. C. BUTLER, B.Sc.,
The University, Reading †.

[Received May 15, 1944.]

ABSTRACT.

A D.C. electrical circuit suitable for the evaluation of certain arithmetical expressions has been devised, and its use for harmonic analysis is described. Apart from a potentiometer and mirror galvanometer, only standard radio components are used, and thus the instrument is inexpensive and simple to construct, while its average error is less than 0.1 per cent. of the largest quantity involved. Modifications of the basic circuit are described.

THE instrument is designed to evaluate quantities of the form $\Sigma p_n q_n$, where the q 's are selected from a discrete set of numbers and the p 's can each have any value between 0 and 1. The instrument is, therefore, very convenient for the analysis of a periodic function in the form

$$f(x) = A_0 + \Sigma A_r \cos rx + \Sigma B_r \sin rx, \quad (1)$$

when $f(x)$ is known for an even number N of equally spaced values of x given by

$$x = \frac{2\pi n}{N} \quad (n=0, 1, 2 \dots \overline{N-1}). \quad (2)$$

For it is possible to evaluate the following products :

$$\left. \begin{aligned} \frac{2}{N} \sum_{n=0}^{N-1} f\left(\frac{2\pi n}{N}\right) \cos\left(\frac{2\pi r n}{N}\right) &= A_r^*, \\ \frac{2}{N} \sum_{n=0}^{N-1} f\left(\frac{2\pi n}{N}\right) \sin\left(\frac{2\pi r n}{N}\right) &= B_r^*, \end{aligned} \right\} (3)$$

since they are of the form $\Sigma p_n q_n$, the p 's being (on a suitable scale) the values of $f\left(\frac{2\pi n}{N}\right)$ and the q 's the possible values of $\sin\left(\frac{2\pi r n}{N}\right)$ and $\cos\left(\frac{2\pi r n}{N}\right)$. A_r^* and B_r^* are related to the Fourier coefficients A_r and B_r by the equations

$$\left. \begin{aligned} A_r^* &= A_r + A_{N-r} + A_{N+r} + A_{2N-r} + A_{2N+r} + \dots \\ B_r^* &= B_r - B_{N-r} + B_{N+r} - B_{2N-r} + B_{2N+r} - \dots \\ A_{N/2}^* &= A_{N/2} + A_{3N/2} + A_{5N/2} + \dots, \\ B_{N/2}^* &= 0. \end{aligned} \right\} r < \frac{N}{2}. \quad . (4)$$

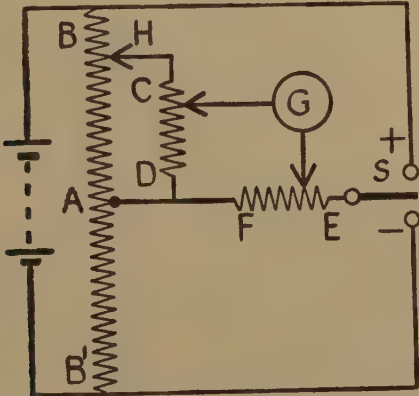
† Communicated by Professor J. A. Crowther, M.A., Sc.D.

Provided no harmonic higher than the $N/2$ th. is present, $A_r^*=A_r$ and $B_r^*=B_r$. If this is not the case, then it is impossible to eliminate the uncertainty in A_r and B_r if $f(x)$ is known only for a limited number of values of x , and therefore its complete functional form is unknown.

The electrical circuits have to perform the operations of multiplication and addition. Fig. 1 shows the principle of the multiplying circuit. BAB' is a centre-tapped potentiometer of low resistance connected across a battery, while CD and EF are high resistance potentiometers. Each potentiometer is calibrated to read the ratio of its output to input voltage, the scales of CD and EF each extending over the range 0 to +1, while that of BAB' extends over the range +1 to -1. Then the galvanometer G will indicate zero current if

$$r=pq, \dots \dots \dots (5)$$

Fig. 1.



Multiplying circuit.

r being the setting of EF, p that of CD and q that of BAB'. The switch S will have to be in the upper (+) or lower (-) position according as the contact H is on the section BA ($1 > q > 0$) or B'A ($0 > q > -1$) of BAB', *i. e.* according as the product pq is positive or negative. Hence the circuit serves to measure the magnitude and sign of the product of two numbers. A necessary condition for the validity of (5) is that the resistances of CD and EF should each be large compared with that of BAB'.

Fig 2 (a) shows the adding circuit. The numbers to be added are represented by e.m.f.'s $E_1, E_2, \dots E_N$. Then it follows from Thévenin's theorem that this circuit is equivalent to that of fig. 2 (b) if

$$E = \frac{E_1 + E_2 + \dots + E_N}{N} \quad R' = R/N. \dots \dots \dots (6)$$

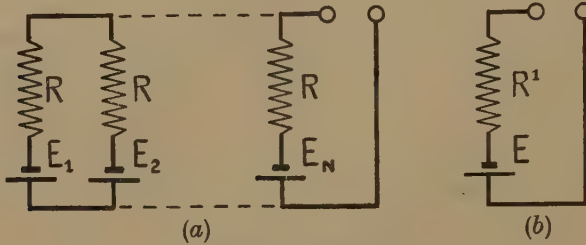
The circuit therefore has the property of averaging a number of potentials.

We will now describe an instrument, based on these principles, for performing an harmonic analysis of the kind described in the first

paragraph. To simplify the description, we shall assume that the instrument is designed to deal with functions known at every 20° of their period, *i. e.* that $N=18$. The modifications necessary for other values of N will be obvious.

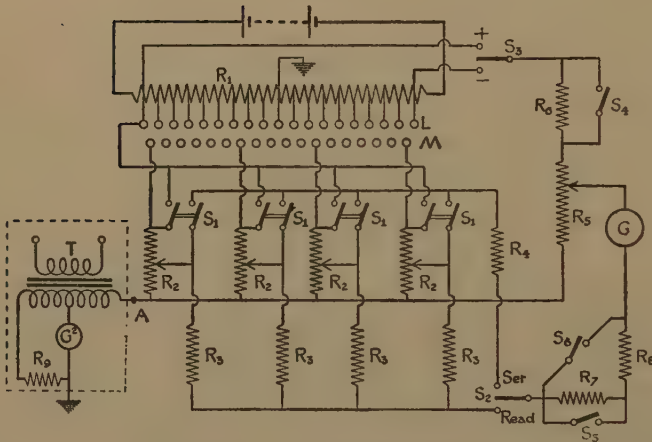
The complete circuit is given in fig. 3. R_1 is a 15-ohm potentiometer wire connected across a 12-volt accumulator. It corresponds to BAB' of fig. 1. Nineteen tappings on R_1 are brought out to a row of mercury cups L . These are numbered from left to right 0, 1, . . . 18. The middle

Fig. 2.



Averaging circuit.

Fig. 3.



Complete circuit of instrument.

cup (No. 9) is connected to earth, and the tappings on R_1 are made so that the potential differences between the cups and earth are proportional to $\cos 0^\circ, \cos 10^\circ, \cos 20^\circ, \dots \cos 180^\circ$, the potential of cup No. n to earth being $\cos 10n^\circ$.

A second row of eighteen mercury cups M is arranged parallel to the first, and each is connected to a separate 10,000-ohm potentiometer R_2 ; the other ends of these are all connected to A . To avoid confusing the diagram, only four of the R_2 's have been drawn. For the present, we will neglect the portion of the circuit on the left which is enclosed in a

dotted rectangle, and imagine that the point A is connected to earth, *i. e.* to the middle tapping of R_1 . The R_2 's correspond with CD of fig. 1 and their scales are each graduated from 0 to 1. Each sliding contact is connected through a separate 1-megohm resistor R_3 to the lower contact of S_2 . The R_3 's correspond with the R 's of fig. 2 (a) and the outputs from the R_2 's correspond with the potentials averaged. S_2 is connected through R_7 , R_8 and the galvanometer G to the moving contact of a 10,000-ohm potentiometer R_5 , which is calibrated from 0 to 1. This corresponds with EF of fig. 1, and is connected between A (or earth) and either the + or - end of R_1 according to the position of S_3 .

Consider all the switches S_1 to be open, S_2 to be on the lower contact (marked *Read*) and S_4 to be closed. The purpose of these switches will be explained later. Then suppose we wish to find the coefficient A_1^* of equation (3). The sliders of the R_2 's are set at the values of $f(x)$ corresponding to the values of x equal to 0° , 20° , 40° , 340° respectively, and the mercury cups L and M are connected so as to apply potentials to the R_2 's equal to $\cos 0^\circ$, $\cos 20^\circ$, $\cos 340^\circ$. This is done by connecting the cups M in order from the left to the following numbered cups L :

0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 16, 14, 12, 10, 8, 6, 4, 2.

For the potential of cup No. n is $\cos 10n^\circ$ and $\cos 200^\circ = \cos 160^\circ$, $\cos 220^\circ = \cos 140^\circ$, etc.

Then it will be seen that the potentials between the sliding contacts of the R_2 's and earth are, in order,

$$f(0^\circ) \cos 0^\circ, f(20^\circ) \cos 20^\circ, \dots \dots f(340^\circ) \cos 340^\circ.$$

These potentials are averaged by the megohm resistors R_3 in accordance with the principle of fig. 2, and therefore the potential between the lower contact of S_2 and earth is

$$\frac{1}{18} \sum_{n=0}^{n=17} f(20n^\circ) \cos 20n^\circ = \frac{1}{2} A_1^*.$$

If, then, R_5 and S_3 be adjusted so that the galvanometer G indicates a balance, the reading of R_5 will be numerically equal to $\frac{1}{2} A_1^*$, and the position of S_3 will indicate the sign of A_1^* .

In order to determine the other coefficients in the expansion of $f(x)$, the R_2 's are left undisturbed and the connections between the cups L and M are changed. For example, it may be seen that, to determine B_1^* , the cups M should be connected in order from the left to the following numbered cups L :

9, 7, 5, 3, 1, 1, 3, 5, 7, 9, 11, 13, 15, 17, 17, 15, 13, 11.

We now proceed to consider the elements of the circuit in more detail. R_5 is a 3-decade, 10,000-ohm potential divider. A decade type of potential divider is particularly convenient because of the ease of reading

it. When the higher coefficients of a Fourier series are being determined, the readings of R_5 will usually be very low, and it is then convenient to be able to increase them by a factor of ten, thus adding an extra figure to the result. This is done by throwing in the 90,000-ohm resistor R_6 by opening the switch S_4 .

The galvanometer G needs to be of short period, or else the operation of the instrument becomes very tedious. A critically damped, taut suspension, mirror galvanometer of period 2 seconds and a sensitivity of 20 mm./ μ A. serves admirably. The zero of such an instrument remains constant to a millimetre for long periods, while the deflection produced by a change of 0.001 in R_5 is about a centimetre. This high sensitivity makes the adjustment of R_5 for balance a very quick operation. R_7 and R_8 are radio resistors of values 5 megohms and 0.5 megohm respectively. Together with the press-button switches S_5 and S_6 , they serve to control the sensitivity of the galvanometer circuit during the early stages of balancing R_5 . It is desirable that S_5 and S_6 should be of press-button type so that the galvanometer circuit is automatically left in the state of low sensitivity while the R_2 's or the mercury cup connectors are being adjusted.

The R_2 's are wire-wound radio potentiometers costing a few shillings each, and prove to be very reliable. They can be fitted with a pointer knob to move over a circular scale, and this can be calibrated. Such radio potentiometers have been found to retain their calibration to better than 2 per cent. after several years' continuous use. Thus, provided that a 2 per cent. accuracy in the ordinates of $f(x)$ is sufficient, they may be set directly. If, however, greater accuracy is desired, they are set by means of the precision potentiometer R_5 . The switch S_2 is changed over to the upper contact (marked *Set*) and any connector present is removed from the mercury cups L and M . Then to set any ordinate, R_5 is first set to the desired value and the switch S_1 associated with the appropriate R_2 is closed. This connects the input of the R_2 to potential $+1$ and the output through R_4 and S_2 to the galvanometer circuit. R_2 is then adjusted until the galvanometer indicates a balance. This can be accomplished to better than 0.002, and all ordinates can be set up to this accuracy in a few minutes. R_4 is a 60,000-ohm radio resistor and serves to give the galvanometer circuit approximately the same sensitivity for the setting operation as for the reading of the Fourier coefficients. It should be noted that if one of the switches S_1 is accidentally closed while a connector is between the mercury cups LM , R_1 may be partially short-circuited and damaged. It is advisable to insert fuses in the battery leads to guard against this possibility.

The R_3 's are 1-megohm radio resistors. They are tested on a Wheatstone bridge and adjusted to equality by filing. It may seem surprising that such resistors are sufficiently constant for a circuit of this kind. It is true that wire-wound resistors are to be preferred. If, however, these are unobtainable, a serviceable instrument can be made using the

composition type. In our own instrument, all the resistors were initially adjusted to be within 0.1 per cent. of their mean value. After a lapse of four months, the average deviation of a single one from the mean value was only 0.15 per cent. It will be shown later that this represents an error of only 0.02 per cent. in the determination of the Fourier coefficients. As these resistors have a high temperature coefficient, it is desirable to mount them as close together as possible and to enclose the whole set in a metal box in order that changes of temperature may affect them all equally.

The contacts between the mercury cups L and M are made by connectors, each consisting of a strip of insulated material fitted with two rows of amalgamated metal pegs. These pegs dip into the mercury cups and are joined together in the appropriate manner. A separate connector of this type is used for each Fourier coefficient. Since there are several hundred pegs in the different connectors of a complete set, it is desirable to have a quick method of assembly. Strips of $\frac{1}{2}$ -in. plywood are used as the insulating medium. Holes are bored with a No. 45 drill and the pegs, which are simply No. 4 brass wood-screws $\frac{3}{4}$ -in. long, are driven in with a ratchet screw-driver so that the heads are flush with one side of the wood and the points project about $\frac{1}{8}$ in. on the other side. The necessary connections between the screws are made by soldering on to the heads. Tinning the points of the screws assists amalgamation. This method of construction is quick and does not require highly skilled labour.

Mercury contacts have been long recognized as a simple means of securing a trouble-free and low resistance contact. Nevertheless, it is desirable, in an instrument of this kind, to have some means of indicating that the connections are all good. This is provided by the portion of the circuit on the left which is enclosed in a dotted rectangle. The point A is not connected directly to earth as has hitherto been supposed. A little consideration will show that, when a connector is placed between L and M, the resistance of the instrument between A and earth is composed of the eighteen R_2 's and R_5 , each in series with a section of R_1 , and connected in parallel. If there is a bad contact at one of the mercury cups, resulting in one of the R_2 's being disconnected, then this combined resistance changes by at least 5 per cent. of its normal value, and this change is detected in the following way: T is a 4-volt transformer with centre-tapped secondary, R_9 a resistance of about 500 ohms (the exact value being found by trial) and G^2 any *linear* detector of A.C. having a low D.C. resistance. We use a vibration galvanometer, but other means (for example, an amplifier and "magic eye") would serve equally well. The fourth arm of the bridge is formed by the resistance of the instrument between A and earth. A faulty contact is indicated by a current in G^2 . Since G responds only to D.C. and G^2 to A.C., the presence of the A.C. test circuit does not disturb the operation of the instrument, neither does the direct current flowing in the main circuits disturb the A.C. balance. It is, however, often convenient to switch off the A.C. circuit

(by a switch in the primary of T) while R_5 is being adjusted to prevent surges in the galvanometer circuit on closing S_5 or S_6 .

R_1 consists of four metres of resistance wire having a resistance of 15 ohms. The tapplings are made as follows. The approximate positions are first marked, using an ordinary potentiometer. At each position, a loop of wire about 6 in. long and having a resistance of about 2 ohms is joined by soldering between two points 1 in. on either side of the approximate position. The tapping connections are then made to points on these loops. By this means, it is possible to adjust one of the tapplings without disturbing the others by making and unmaking soldered joints on the main wire. In this way, it is easy to place all the tapplings correct to considerably better than 0.001, the length of either half of R_1 being taken as unity.

There are the following sources of error in the instrument: errors of setting R_2 , errors in the initial adjustment of R_1 and R_3 , loading of the potentiometers R_2 by R_3 , loading of the main potentiometer R_1 by the R_2 's.

It can be shown that the probable error in any harmonic due to the setting of R_2 is $\sqrt{2/N}$ of the probable error of setting each R_2 . Since these can be set to better than 0.002, it follows that for $N=18$, the probable error in any harmonic due to this cause is less than 0.0007.

Errors in the initial adjustment of R_1 will have a similar effect to errors in the setting of the R_2 's and will contribute less than 0.0001 to the errors of any harmonic.

Errors due to inequalities of the megohm averaging resistors will result in unequal weights being given to the outputs of the different R_2 's. If the probable error of a single averaging resistor is r , the probable

error of a single harmonic will be $\frac{r}{R_3} \sqrt{\frac{1}{3N}}$. The error r arises from errors in the megohm resistors themselves, amounting to about 1500 ohms, and changes in the output resistance of R_2 , which is in series with R_3 . The output resistance of R_2 varies between 0 and 2500 ohms, depending on the setting, its average value being 1667 ohms. The probable error in any harmonic from both sources is thus about 0.0003.

The magnitude of the error due to loading of R_1 depends on the exact way in which the R_2 's are distributed along R_1 , and is therefore different for different harmonics. Detailed calculations show that if the total resistance of R_1 is 15 ohms, and that of R_2 10,000 ohms, the maximum possible error from this cause is 0.0004, while the average error is much less.

We should therefore expect that the average error in any harmonic due to all causes would be of the order of 0.001. Experimentally, the accuracy of the instrument has been tested by using it to analyse saw-tooth waves of the form

$$f(x) = \frac{a-x}{360} \quad 0 < x < a,$$

$$f(x) = 1 + \frac{a-x}{360} \quad a < x < 360,$$

x and α being measured in degrees. The Fourier coefficients of such waves can easily be calculated. It was found that the average error in a coefficient was 0.0007. Since this average error is independent of the magnitude of the coefficients, the *percentage error* is smallest when the scale of $f(x)$ is so chosen that its variation extends over the full range of the instrument from 0 to +1. The magnitude of the average error shows that when R_8 is introduced in order to multiply the reading of R_5 by ten, the last figure cannot claim high significance. Its retention, however, prevents "rounding-off" errors accumulating.

As was indicated in the first paragraph, additional error arises from the fact that $f(x)$ is known only at a number of discrete values of x . It follows that, unless the Fourier series converges fairly rapidly, the errors due to the presence of high harmonics will far outweigh those due to the electrical circuits. The first step towards greater accuracy is therefore to increase N , the number of points in the period at which $f(x)$ is known. One way of doing this is, of course, to increase the number of the R_2 potentiometers and of the tappings on R_1 . It is, however, possible, at the expense of rapidity of operation, to perform an analysis of a function with N ordinates on an instrument with only $\frac{1}{2}N$ potentiometers, *e. g.* to use the instrument described above to analyse a function known at every 10° .

We have from (3)

$$\left. \begin{aligned} A_r^* &= \frac{2}{N} \sum_{n=0}^{n=\frac{1}{2}N-1} f\left(\frac{2\pi n}{N}\right) \cos\left(\frac{2\pi r n}{N}\right) \pm \frac{2}{N} \sum_{n=\frac{1}{2}N}^{n=N-1} f\left(\frac{2\pi n}{N}\right) \cos\left(\frac{2\pi r(n-\frac{1}{2}N)}{N}\right), \\ B_r^* &= \frac{2}{N} \sum_{n=0}^{n=\frac{1}{2}N-1} f\left(\frac{2\pi n}{N}\right) \sin\left(\frac{2\pi r n}{N}\right) \pm \frac{2}{N} \sum_{n=\frac{1}{2}N}^{n=N-1} f\left(\frac{2\pi n}{N}\right) \sin\left(\frac{2\pi r(n-\frac{1}{2}N)}{N}\right), \\ A_0^* &= \frac{1}{N} \sum_{n=0}^{n=\frac{1}{2}N-1} f\left(\frac{2\pi n}{N}\right) + \frac{1}{N} \sum_{n=\frac{1}{2}N}^{n=N-1} f\left(\frac{2\pi n}{N}\right), \end{aligned} \right\} \quad (7)$$

the positive or negative sign being taken according as r is even or odd. Connectors can be constructed such that the first terms in the expressions for A_r^* and B_r^* are the readings given by an instrument with $\frac{1}{2}N$ potentiometers when these are set at the values of $f(x)$ occurring in the first half period, while the second terms are the readings when the potentiometers are set at the values of $f(x)$ in the second half period. The procedure is, therefore, to set up on the instrument the values of $f(x)$ in the first half period and then, by inserting the different connectors in turn, to determine quantities a_r and b_r which are the first terms of equations (7). The values of $f(x)$ in the second half period are then set up and the readings are repeated, giving a'_r and b'_r . Then

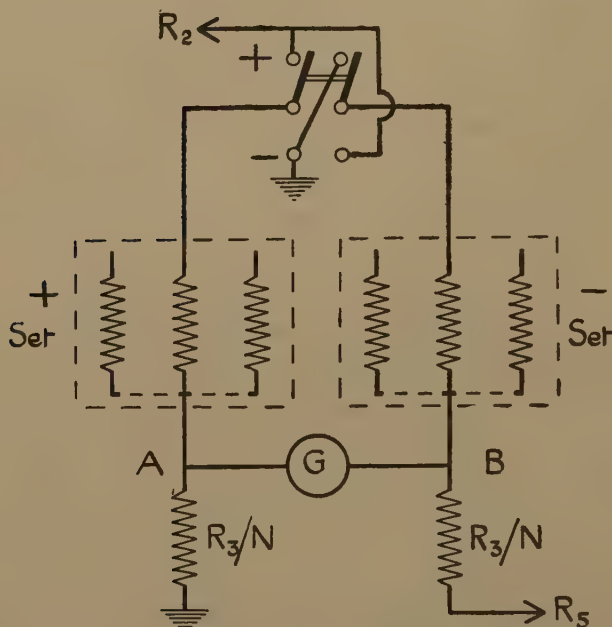
$$\begin{aligned} A_r^* &= a_r \pm a'_r, \\ B_r^* &= b_r \pm b'_r, \\ A_0^* &= \frac{1}{2}(a_0 + a'_0). \end{aligned}$$

This is the way we regularly use our instrument, which has eighteen

potentiometers, to analyse functions at every 10° . We find that about two hours are required to determine the complete set of $36 A_r^*$'s and B_r^* 's within the limits of accuracy mentioned above; it is clear that this time would be approximately halved by the use of an instrument with 36 potentiometers. It may be noted that the connector for the r th harmonic when the instrument is used with $\frac{1}{2}N$ ordinates is the same as that for the $2r$ th harmonic when it is used with N ordinates.

It is clear that this instrument will not only perform harmonic *analysis* but also serves for *synthesis* according to equation (1). The N quantities A_r^* and B_r^* obtained from the Fourier coefficients according to equation

Fig. 4.



Circuit for introduction of negative amplitudes.

(4) are set up on the N potentiometers of the instrument, and a special set of N connectors is used to give the values of $f(x)$ for the N values of x of equation (2). In using the instrument in this way, a difficulty arises owing to the fact that only *positive* numbers can be set on the potentiometers, while some of the amplitudes may be negative. An analogous difficulty does not arise when the instrument is used for analysis because it is possible to add to the function being analysed a constant term large enough to make all the ordinates positive; this added term is then subtracted from the reading A_0^* given by the instrument.

There are two ways of surmounting this difficulty. One way, which involves no modification in the circuit, consists in setting up on the instrument the *positive* A_r^* 's and B_r^* 's and setting the remaining potentiometers to zero.

meters to zero. The readings of the machine then give a function $f_+(x)$ which is the synthesis of all positive harmonics. The negative A_r^* 's and B_r^* 's are then set up in the same way, and the readings of the instrument then give a function $f_-(x)$ which is the synthesis of all the negative harmonics. The required function is then clearly

$$f(x) = f_+(x) - f_-(x).$$

The second method is quicker in operation, but requires a rather more elaborate circuit. Two separate sets of averaging resistors R_3 are provided. The outputs of all the potentiometers R_2 for which a positive sign is required are connected to the R_3 's of one set and those for which a negative sign is required to the R_3 's of the other set. The R_3 's which are not connected to an R_2 are earthed. Fig. 4 shows how a double-pole change-over switch may be used to connect the output of an R_2 to the resistor of one set and to earth the corresponding resistor of the other set. The reading which is required is then the difference of the potentials E^+ and E^- produced by the two sets. This may be measured by R_5 in the following way. The output of the $+$ set of R_3 is connected to earth through a resistance R_3/N . Reference to fig. 2 will show that the potential of the point A to earth is then $\frac{1}{2}E^+$. The output of the $-$ set of R_3 is connected through a resistance R_3/N to the output of R_5 . If the reading of R_5 be E , it follows from the principle of fig. 2 that the potential to earth of the point B is $\frac{1}{2}(E + E^-)$. If the galvanometer circuit connected between A and B indicates a balance, we must have

$$\frac{1}{2}(E + E^-) = \frac{1}{2}E^+.$$

Hence

$$E = E^+ - E^-.$$

This arrangement requires that the output resistance of R_5 is either constant or negligible compared with R_3/N . If a low resistance potentiometer be used for R_5 , it will be necessary, in order to prevent a disturbance of the potential distribution along R_1 , to connect a balancing resistance across the half of R_1 opposite to R_5 . The necessary modifications in the switches S_2 and S_3 are easily worked out.

For many purposes high precision is unnecessary. If an average error of 1 per cent. can be tolerated, the instrument can be greatly simplified. The circuits for setting the R_2 's can be omitted, and R_5 can be replaced by a radio potentiometer similar to those used for R_2 . Alternatively, the galvanometer and R_5 can be omitted and a centre-zero microammeter reading from -100 to $+100 \mu A$. can be connected between the output of the R_3 's and earth (if the instrument is to be used only for positive quantities) or between the outputs of the $+$ and $-$ sets of R_3 's (if the instrument is to be used for both positive and negative quantities). Some form of standardizing circuit would be needed so that the current through R_1 could be adjusted to the correct value. Such an instrument would be relatively inexpensive and simple to construct, and would appear to be very suitable for work in X-ray crystallography.

As an example of the versatility of an instrument of this type, it may be mentioned that we have had occasion to evaluate quantities of the form $\Sigma na_n / \Sigma a_n$. The instrument gives the numerator and denominator if the usual potentiometer wire R_1 is replaced by one with equally spaced tappings.

Acknowledgment.

It is with pleasure that we record our appreciation of the interest taken in this work and the facilities placed at our disposal by Professor J. A. Crowther.

LXX. *On the Dimensions of Physical Magnitudes* (Fifth Paper).

By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.*

[Received September 19, 1944.]

THE fundamental difference between my view of dimensions and that of Professor Wilson is simply this. I use the dimensional symbol as standing for a particular operation of measurement, the result of which yields a numerical magnitude. If we find by experiment that the magnitudes obtained by performing the distinct operations, X and Y, are proportional to one another, the constant of proportionality must be dimensional; thus, if we find $X = \alpha Y$, α must have the dimensions $[XY^{-1}]$. Professor Wilson, on the other hand, believes that there are things called "physical quantities," which we may be able to measure by various methods, and that each physical quantity has some characteristic called its "dimensions," our statement of which is "correct" or "incorrect" quite independently of the operation of measurement. We may even be unable to arrive at any idea of them; nevertheless, the correct value still exists and may one day be found. I use the "method of dimensions" to represent what we do, and so to enable us to distinguish clearly between what we learn from experiment and what is implied by our definitions. In Professor Wilson's view, dimensions are simply one of many characteristics of an objective world which we try to discover as best we can.

All this, I think, has already been set out clearly enough for readers to make their own choice, and I should not have written further on the point had not Professor Wilson in his latest paper † persisted in reading into my words his own preconceptions and foisting on me the nonsense which naturally results. His or my fundamental view may be either right or wrong, but if he wishes to criticize my development of my own premises, he should not in doing so assume that it is intended to be a development of his.

* Communicated by the Author.

† Phil. Mag. xxxv. p. 420 (1944).

For example, he quotes my equation * $[V]=[\alpha E']$, and because, according to *his* meaning of V and E' , they "can be, and frequently or usually are, measured by a voltmeter," he deduces that their dimensions, whatever they are, must be identical, so that α must be dimensionless. But in my paper I defined the operations of measurement which I represented by V and E' , and they are different. Hence the proportionality of the corresponding magnitudes is an experimental result, and so α must have dimensions. Professor Wilson, taking his own view of things, may believe that the "correct" dimensions of V and E' are identical—that is another matter—but he has no right to say that I must reach that conclusion.

Professor Wilson now says that, when he deduced † $[M]=[T^3]$ from my premises he realized that he had reached an absurdity, and, in fact, started from those premises in order to show that they would lead to it. I fear he must have written this from memory. If he had referred back to his argument † he would have found that this explanation was quite inadmissible. If it were the true explanation he would have stopped at the relation in question and merely added "which is absurd." In fact, however, he goes on, without pause, to deduce the "correct" dimensions for $\mu\kappa$. The facts of the matter are these. He wanted to show that, *even from my premises*, the "correct" result would follow if only one of my equations were amended. ("It is some relief to me," he wrote, "to find that, on replacing the erroneous statement (14) by the correct one, (15), the less startling result, $[\mu^{-1}\kappa^{-1}]=[LT^{-1}]$, emerges"—this result, in fact, being one which he regards as "correct.") He "amended" the equation, and the "absurdity" in question then appeared in the course of the deduction of the "correct" result. If he had recognized it as an absurdity his whole argument would have collapsed, and he could not possibly have experienced the relief which he tells us descended upon him.

In his latest argument on this point, Wilson writes, "According to Dingle, it [the constant β] has the same dimensions as a dielectric constant." I wish that Professor Wilson, even if he cannot understand what I write, would at least refrain from inventing statements for me as though I were a character in a play he was writing. I have never made or implied this meaningless statement, and even the most casual reading of my paper would have shown that while, in the scheme I proposed, α and β performed the same function as μ and κ in the current system, there was no one-to-one correspondence between them; it was, in fact, to correct the contrary implication of my former paper that I wrote it. "It is," I wrote, "only in a very general sense that α and β can be said to be equivalent to μ and κ ." In pointing out the relation between the two schemes I said, "whereas in the current scheme the dimensions of μ and κ are 'unknown' . . . the dimensions of α and β

* Phil Mag. xxxiv. p. 594 (1943).

† Phil. Mag. xxxiii. p. 635 (1942).

are determined by the manner in which they appear in the equations.” If Professor Wilson had merely asserted that β had the same dimensions as κ , it would have shown that he had not read my paper before criticizing it. To attribute the assertion to me is even less pardonable. It does, however, absolve me from commenting on the irrelevant matter which follows.

Wilson’s historical discussions are interesting, but quite beside the point. He asks for the grounds on which I base the “dogmatic pronouncement” that it is “literally nonsense” to speak of a mass of so many metres. I should have thought the grounds were obvious; mass is measured canonically by a kind of operation which does not give a result in metres. I was, of course, not asserting, as the context shows clearly enough, that those who so speak are speaking nonsensically. The “operative word,” to use modern jargon, was “literally.” Milton’s “darkness visible” is literally nonsense, and if one is discussing the literal meaning of words it may be necessary to say so. It does not follow that one would prefer “insipissated gloom.”

In the matter of the “fine-structure constant,” I was discussing Eddington’s statement that $\frac{ch}{2\pi e^2}$, to which he gives this name, is dimensionless; I gave the reference to Eddington’s book. I pointed out that this quantity was not dimensionless, but that the quantity $\frac{ch\kappa}{2\pi e^2}$, according to the current scheme, or $\frac{ch\beta}{2\pi e^2}$, according to my own, was so. I went on to discuss the significance of the additional factor. Professor Wilson now makes the charge, “Professor Dingle is therefore (no doubt inadvertently) misrepresenting the fine-structure constant when he assumes that it does not contain κ .” What can I say?

It would give me great pleasure if Professor Wilson would try the experiment of temporarily putting out of his mind the views on dimensions which he has held for so long, and reading my papers *ab initio*. He would then see that what he calls my logical errors are merely incompatibilities between my conception of the subject of dimensions and his own. Indeed, it is scarcely possible for me to make a logical error in this matter, because the procedure is so extremely simple. I have only to ask myself what operations of measurement I am talking about, and then write down the corresponding symbols. I may, of course, have failed in selecting suitable processes of measurement, or in expressing what I want to say unambiguously, but I should have to be very drowsy indeed to make such mistakes as Wilson ascribes to me, which are so many and various as to suggest the zeal of the collector rather than the perplexity of a constructive critic. Professor Wilson has shown his readiness to adopt my premises “for the sake of argument.” Will he not now, for the sake of stopping argument, venture to assume that no one still at large could be quite such a fool as I appear, and see what follows?

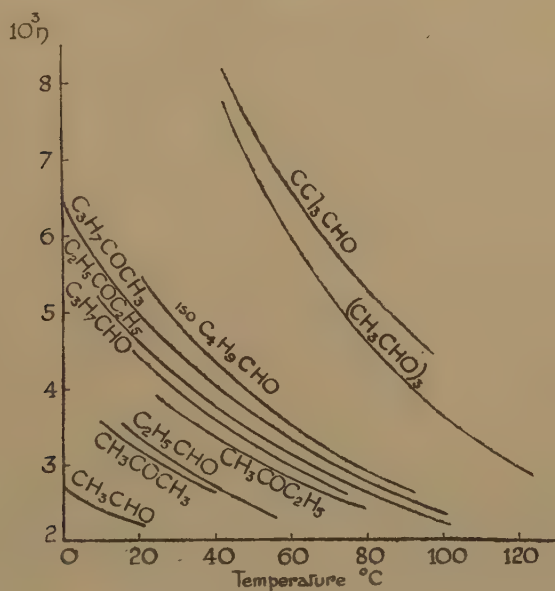
LXXI. Viscosities and Rheochors of Aldehydes, Nitriles and of Secondary and Tertiary Amines.

By J. NEWTON FRIEND and WILLIAM D. HARGREAVES*.

[Received April 21, 1944.]

THE objects of the present research have been to study the viscosities up to the boiling point of some aldehydes, nitriles, secondary and tertiary amines, as well as to determine the rheochors of CN, NH and

Fig. 1.



Viscosity-temperature curves of aldehydes and isomeric ketones.

N(—) respectively. Some primary amines have already been studied ⁽¹⁾. The rheochor has been defined ⁽²⁾ as

$$R = M(10^3 \eta)^{1/8} / (D + 2d),$$

where M is the molar weight, D and d are the density of liquid and vapour respectively, and η the viscosity. In evaluating the present

* Communicated by the Authors.

results use has been made of the following rheochors, which have been established in previous papers.

C	12.8	Cl	27.3
H (in C—H)	5.5	NH ₂	20.6
H (in O—H)	10.0	C ₆ H ₅	
O<(etheric)	10.0	Attached to alkyls	100.7
O=(ketonic)	13.2	Attached to other groups	102.7
Covalent bond	0.0	Saturated six - membered	
Co-ordinate bond... ..	-0.4	carbon ring.....	-5.6

To these may be added CN, 33.0 ; NH, 13.6 ; and N(—), 6.6 ; established in this paper.

Aldehydes.

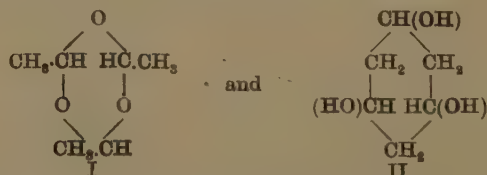
The viscosities of the aldehydes are seen in fig. 1 to increase steadily with the molecular weight. Their boiling-point data are summarized

TABLE I.
Boiling-point Data for Aldehydes.

	10 ³ η.	D.	R (obs.).	R (calc.).
CH ₃ CHO.....	2.21	0.7755	62.4	60.8
(CH ₃ CHO) ₃	2.83	8672	172.0	172.0
CCl ₃ CHO	4.47	1.3718	128.6	126.2
C ₂ H ₅ CHO	2.54	0.7693	84.3	84.6
<i>n</i> C ₃ H ₇ CHO	2.64	7455	108.5	108.4
<i>iso</i> C ₄ H ₉ CHO	2.61	7273	132.5	132.2

in Table I. As usual, the first member of the series, viz., acetaldehyde, is slightly irregular, giving a rheochor 1.6 units too high. Chloral also gives a high result, due, presumably, to the mutually repulsive effect of the three chlorine atoms attached to the same carbon, for both CCl₄ and CBr₄ have been shown to yield high rheochors⁽²⁾.

The rheochor of paraldehyde is of interest. As mentioned by Glassstone⁽³⁾, the parachor, using Sugden's data⁽⁴⁾, does not enable a decision to be made between the two possible cyclic formulæ—



as no distinction is made between H attached to C and H attached to O in hydroxyl.

The observed rheochor of paraldehyde is 172.0; that of acetaldehyde is 62.4. If formula I. is correct, the differences between paraldehyde and 3 molecules of acetaldehyde consist in the removal of three ketonic double bonds and formation of a saturated 6-membered ring. Whence the rheochor of paraldehyde should be $3 \times 62.4 - 3 \times 3.2 = 5.6$ or 172.0. For formula II. the rheochor is 185.5, so the evidence is strongly in favour of formula I. The actual identity of the observed and calculated values for formula I. is, however, fortuitous. It is assumed that the rheochor of the 6-membered ring is the same as for a saturated 6-carbon ring. This is probably not quite correct, although the difference can only be small.

The viscosities of the aldehydes, as shown in fig. 1, are greater than those of their isomeric ketones⁽⁵⁾ at the same temperatures. On the other hand, the boiling points of the aldehydes are lower. Molecular symmetry would appear to be an important factor here, the more symmetrical molecules having lower viscosities than their less symmetrical isomerides. This is well illustrated by the steady rise in viscosity on passing successively from di-ethyl ketone through methyl propyl ketone to the even less symmetrical *iso*-valeric aldehyde.

The densities of the aldehydes are also greater than those of the ketones and the rheochor does not enable a distinction to be made; the CO group in aldehydes thus yields the same rheochor as in ketones.

Nitriles.

Viscosity-temperature curves for the nitriles are shown in fig. 2, the data for HCN being those of Fredenhagen⁽⁶⁾. As usual, the effect of the first CH_3 is much greater than that of successive replacements; this is no doubt connected with the disappearance of the linear polymerization of HCN due to the replacement of the one H atom that can exert a fairly powerful hydrogen-bond effect. Acetonitrile and the higher alkyl nitriles cannot very well form hydrogen bonds⁽⁷⁾.

The boiling-point data for the nitriles are given in Table II.

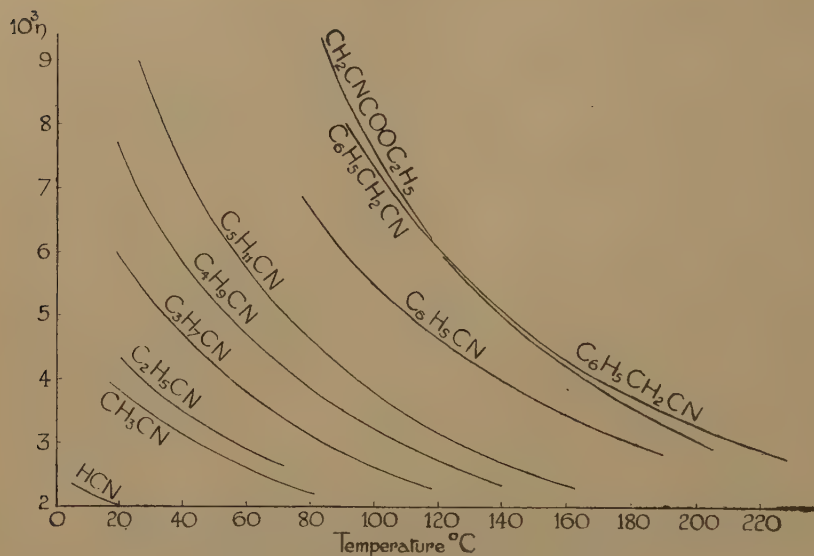
The two aromatic nitriles illustrate the general rule that substitution in the side chain has less influence on the rheochor of the C_6H_5 group than replacement of a nuclear hydrogen⁽¹⁾ by the same group.

Amines.

It has already been pointed out⁽⁸⁾ that the replacement of hydrogen in the hydroxyl group of an alcohol or phenol appreciably reduces the viscosity, despite the increase in molecular weight. This is attributed in general to elimination of the hydrogen bond. The curves in fig. 3 show that successive replacements of hydrogen in the amino group of aniline produce similar effects, and presumably for the same reason. Thus dimethyl aniline has an even lower viscosity than methyl aniline, which latter has still one hydrogen in the amino group capable of exerting a bond. The curves also show that, when once the hydrogen bonds have been eliminated, further increase in weight of the substituent

alkyl group raises the viscosity in the normal manner. The enhanced viscosity consequent upon introduction of the phenyl group is clearly

Fig. 2.



Viscosity-temperature curves of nitriles.

TABLE II.
Boiling-point Data for Nitriles.

	$10^3 \eta$.	D.	R.	
HCN	2.01	0.6953	42.3	R(CN). 33.7 32.8 32.7 32.6 32.5 34.5
CH ₃ CN	2.22	0.7168	63.0	
<i>n</i> C ₂ H ₅ CN	2.19	7037	85.9	
<i>n</i> C ₃ H ₇ CN	2.34	6969	109.6	
<i>n</i> C ₄ H ₉ CN	2.40	6911	133.3	
<i>n</i> C ₅ H ₁₁ CN	2.36	6838	157.0	
CH ₂ CN.COOC ₂ H ₅ .	2.91	8713	147.4	
Weighted Mean.....				33.0
C ₆ H ₅ CN	2.85	0.8493	138.0	R(C ₆ H ₅). 105.0 101.5
C ₆ H ₅ CH ₂ CN	2.76	8347	158.3	

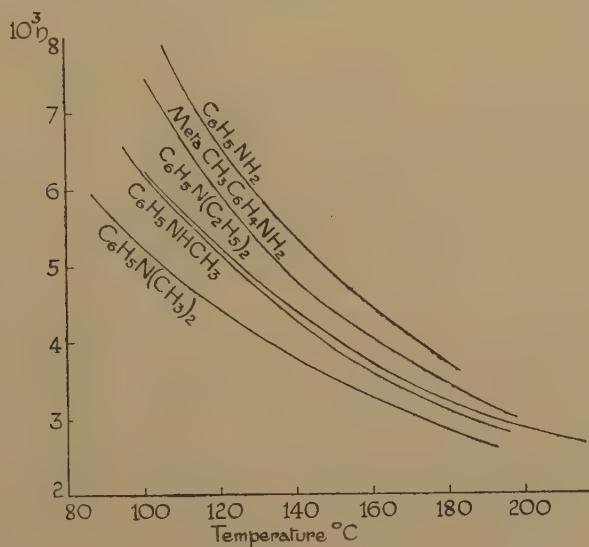
shown in both figs. 4 and 5. The curves for dimethyl aniline and dimethyl paratoluidine are virtually coincident,

The boiling-point data are summarized in Table III.; in evaluating the rheochors of NH and N, the mean value $R=102.7$ has been used for C_6H_5 . The rheochor of NH_2 has already been determined as 20.6.

In evaluating $R(NH)$ from piperidine, in the absence of other data, the ring effect is taken as before as -5.6 , which is the value found for a saturated 6-membered carbon ring; it is not likely to be very different.

It has already been shown⁽²⁾ that whilst the rheochor of hydrogen attached to carbon is 5.5, its apparent rheochor when combined with other elements may be considerably greater; thus in OH it is apparently 10. The following data suggest that, when attached to nitrogen, the

Fig. 3.



Viscosity-temperature curves of primary, secondary and tertiary aromatic amines.

value is approximately 7. The rheochor of ammonia has been calculated using the mean of Elsey's figure⁽¹⁰⁾ of $10^3\eta=2.60$ and Fitzgerald's value, 2.54; this agrees reasonably well with the figure obtained by extrapolation to the boiling point, using Sheppard's formula, from the more recent data of Plank and Hunt⁽¹¹⁾ at higher temperatures.

	R.	R(H).
NH_3	28.3	7.7
NH_2	20.6	7.0
NH	13.6	7.0
N(—)	6.6	

The value for the nitrogen atom refers to it when present as an electro-negative portion of a molecule; when it is positive, as in NOCl, the apparent value is much greater.

TABLE III.

Boiling-point Data for Secondary and Tertiary Amines.

	$10^3\eta$.	D.	R (obs.).	
Secondary amines	R(NH)
$(C_2H_5)_2NH$	2.25	0.6710	119.6	13.4
$n(C_3H_7)_2NH$	2.12	6556	167.8	14.0
$n(C_4H_9)_2NH$	2.124	6441	218.0	16.6
$C_6H_5NH.CH_3$	2.78	0.8400	144.0	12.0
C_2H_5	2.44	7950	169.0	13.2
C_3H_7	2.48	7771	193.2	13.6
C_4H_9	2.38	7556	218.0	14.6
$C_5H_{10}NH$ (piperidine)	4.46	0.7955	128.2	14.8
Weighted Mean....				13.6
Tertiary amines	R(N)
$(C_2H_5)_3N$	2.03	0.6650	166.0	6.7.
$C_6H_5N(CH_3)_2$	2.60	8094	167.4	6.1
$C_6H_5N(CH_3)C_2H_5$	2.44	7821	191.5	6.4
$C_6H_5N(C_2H_5)_2$	2.65	7813	213.7	4.8
$CH_3C_6H_4N(CH_3)_2$ ortho.	2.60	7858	192.1	7.0
para	2.34	7814	190.8	5.7
$CH_3C_6H_4N(C_2H_5)_2$ meta ...	2.41	7516	239.9	7.2
Weighted Mean				6.6

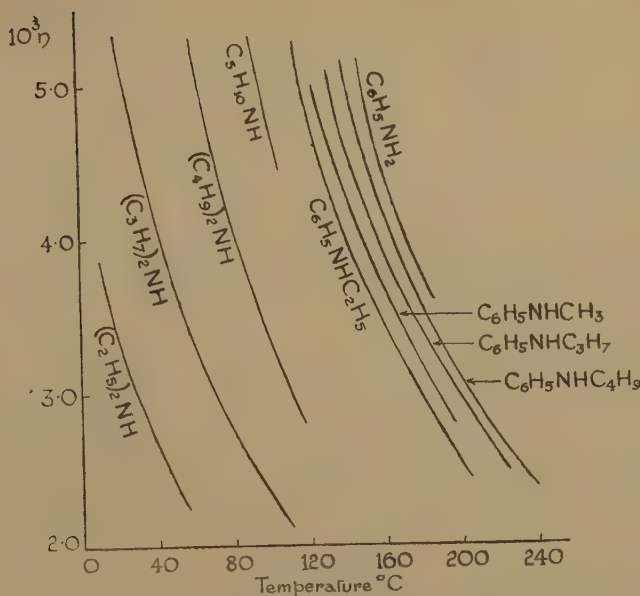
Experimental.

Viscosities were determined with Standard viscometers of the Ostwald type in the manner already outlined ⁽²⁾.

Aldehydes.—These were obtained from the British Drug Houses, dried and fractionated.

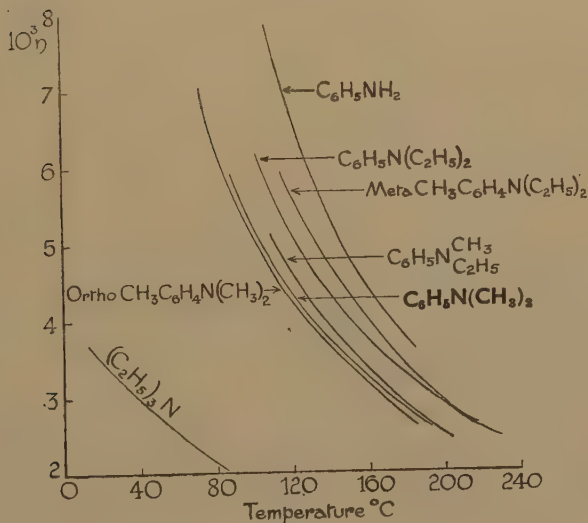
Nitriles.—Acetonitrile was obtained from the B.D.H. Propionitrile was prepared from potassium ethyl sulphate and cyanide ⁽¹²⁾. The higher homologues were made from the corresponding alkyl bromides and sodium cyanide ⁽¹³⁾. All were dried with phosphorus pentoxide and fractionated, the constant boiling point fractions alone being taken..

Fig. 4.



Viscosity-temperature curves of secondary amines.

Fig. 5.



Viscosity-temperature curves of tertiary amines.

Amines.—These were obtained from the B.D.H. dried with potassium carbonate and fractionated.

The viscosity and density data are given below. As before, in all cases, the densities refer to water at 4° C. The values for A and B refer

to Sheppard's equation, $\log_{10} 10^3\eta = A + B/T$, and apply to the linear part of the curve on plotting $\log 10^3\eta$ against $1/T$. The boiling-point values for $10^3\eta$ and D , obtained by extrapolation, are given in parentheses. No correction has been applied for the kinetic energy of efflux as this was extremely small.

Acetaldehyde.—B. pt. 21.6°C . at 755.7 mm. $A = -0.8132$; $B = 341.2$, from 8° upwards.

Temp.	0.1	4.4	8.2	12.2	16.2	20.3	21.6
D	0.8048	0.7990	0.7936	0.7885	0.7829	0.7773	(0.7755)
1000η	2.711	2.600	2.509	2.414	2.326	2.237	(2.213)

$R = 62.4$; calculated 60.8.

The above determinations were made in view of the differences in the densities of the specimens used by Thorpe and Rodger⁽⁹⁾ and by Maass and Boomer⁽¹⁴⁾. Our viscosities lie close to those of the former authors, but our densities are identical with the mean values given by Maass and Boomer and those calculated from the equation of Thorpe and Rodger.

Paraldehyde.—B. pt. 124.1°C . at 751.5 mm. $A = -1.2276$; $B = 666.7$, from 50° upwards.

Temp.	15.0	42.6	51.3	60.1	70.0
D	0.9962	0.9654	0.9558	0.9454	0.9342
1000η	13.24	7.75	6.747	5.957	5.206

Temp.	79.2	91.2	103.9	117.2	124.1
D	0.9233	0.9088	0.8626	0.8759	(0.8672)
1000η	4.642	4.018	3.495	3.027	(2.832)

$R = 172.0$; calculated 172.0.

Our viscosity data agree closely with the two data of Muchin⁽¹⁵⁾ at 15° and 20° respectively.

Chloral.—B. pt. 97.2°C . at 731.3 mm. $A = -0.8448$; $B = 553.4$, from 40° upwards.

Temp.	22.5	41.7	54.6	62.0	70.2	81.2	89.0	94.8	97.2
D....	1.5010	1.4681	1.4455	1.4325	1.4184	1.3993	1.3860	1.3758	(1.3718)
1000η	10.77	8.192	6.966	6.412	5.852	5.225	4.830	4.566	(4.467)

$R = 128.6$; calculated 126.2.

Our results are lower than those of Efremov⁽¹⁶⁾.

Propionaldehyde.—B. pt. 47.8°C . at 751.5 mm. $A = -0.9527$; $B = 435.5$, from 25° upwards.

Temp.	15.4	26.7	31.2	35.35	40.0	44.4	47.8
D	0.8080	0.7950	0.7898	0.7849	0.7795	0.7738	(0.7693)
1000η	3.568	3.167	3.006	2.885	2.742	2.628	(2.540)

$R = 84.3$; calculated 84.6.

Published viscosities are discordant. At 25° , Creighton⁽¹⁷⁾ finds $1000\eta = 4.35$, but Walden⁽¹⁸⁾ gives 3.44. Our interpolated value is 3.23.

n-Butyraldehyde.—B. pt. 74.7 to 74.9° C. at 759.7 mm. $A = -0.8280$; $B = 434.6$, from 30° upwards.

Temp.	18.35	38.7	50.8	60.35	72.2	74.9
D	0.8063	0.7847	0.7715	0.7611	0.7484	(0.7455)
1000 η	4.575	3.673	3.267	2.989	2.696	(2.64)

$R = 108.5$; calculated 108.4.

Iso-Valeraldehyde.—B. pt. 92.5° C. at 757.4 mm. $A = -0.9127$; $B = 486.0$, over the range studied.

Temp.	21.1	35.0	45.3	55.4	65.0	74.7	85.0	92.5
D	0.7976	0.7842	0.7740	0.7641	0.7545	0.7449	0.7347	(0.7273)
1000 η	5.49	4.615	4.113	3.693	3.357	3.063	2.798	(2.61)

$R = 132.5$; calculated 132.2.

Acetonitrile.—B. pt. 81.5° C. at 750.7 mm. $A = -0.7752$; $B = 397.4$, over the range studied.

Temp.	16.6	41.4	51.6	61.8	68.7	76.8	81.5
D	0.7884	0.7615	0.7503	0.7393	0.7318	0.7223	(0.7168)
1000 η	3.95	3.08	2.816	2.583	2.445	2.295	(2.217)

$R = 63.0$; calculated 62.3.

Our viscosity data are slightly lower than those of Walden and Birr⁽¹⁹⁾ (0 to 50°), but our values calculated for 16.1° and 25°, namely 3.98 and 3.62, are close to those given by Muchin⁽²⁰⁾ (3.91) and by Sachanov⁽²¹⁾ (3.60) respectively.

n-Propionitrile.—B. pt. 96.7 to 97.1° C. at 735.9 mm. $A = -0.7809$; $B = 414.8$, from over the range studied.

Temp.	20.1	31.2	43.9	57.4	69.5	83.0	93.1	97.1
D	0.7832	0.7725	0.7597	0.7456	0.7332	0.7189	0.7080	(0.7037)
1000 η	4.305	3.828	3.373	2.982	2.694	2.423	2.250	(2.187)

$R = 85.9$; calculated 86.1.

n-Butyronitrile.—B. pt. 117.8 to 118.0° C. at 740.7 mm. $A = -0.8531$; $B = 477.6$, from 40° upwards.

Temp.	18.0	40.6	59.4	69.9	79.4
D	0.7923	0.7712	0.7535	0.7436	0.7347
1000 η	6.028	4.678	3.832	3.467	3.177

Temp.	90.1	100.3	112.5	118.0
D	0.7242	0.7140	0.7022	(0.6969)
1000 η	2.901	2.667	2.430	(2.34)

$R = 109.6$; calculated 109.9.

n-Valeronitrile.—B. pt. 139.5 to 140.0° C. at 756.4 mm. $A = -0.8428$; $B = 504.9$, over the range studied.

Temp.	18.6	53.8	74.0	86.0	94.4
D	0.7986	0.7687	0.7505	0.7401	0.7329
1000 η	7.746	5.044	4.114	3.675	3.405
Temp.	105.0	114.1	124.2	133.6	140.0
D	0.7229	0.7158	0.7059	0.6971	(0.6911)
1000 η	3.118	2.908	2.679	2.508	(2.40)

$R = 133.3$; calculated 133.7.

n-Capronitrile.—B. pt. 162.2 to 162.7° C. at 745.3 mm. $A = -0.8804$;
B=545.95, from 60° upwards.

Temp.	16.9	57.7	79.3	98.0	108.5	118.1
D	0.8111	0.7761	0.7582	0.7410	0.7319	0.7235
1000 η	10.33	5.908	4.662	3.898	3.551	3.270

Temp.	128.0	137.8	148.1	156.7	162.7
D	0.7148	0.7061	0.6969	0.6892	(0.6838)
1000 η	3.030	2.812	2.610	2.455	(2.359)

R=157.0 ; calculated 157.5.

Ethyl cyanacetate.—B. pt. 207.0° C. at 757.9 mm. $A = -0.9732$;
B=690.0, from 100° upwards.

Temp.	79.7	111.0	127.2	141.6	156.0
D	1.0025	0.9696	0.9536	0.9391	0.9245
1000 η	9.71	6.667	5.643	4.907	4.332

Temp.	171.0	185.8	200.6	207.0
D	0.9090	0.8937	0.8781	(0.8713)
1000 η	3.812	3.389	3.049	(2.91)

R=147.4 ; calculated 145.9.

Benzonitrile.—B. pt. 190.0 to 190.5° C. at 757.9 mm. $A = -0.7215$;
B=545.1, from 75° upwards.

Temp.	19.7	75.0	100.7	120.0	139.9	160.4	180.7	190.5
D	1.0065	0.9573	0.9346	0.9175	0.8992	0.8796	0.8590	(0.8493)
1000 η	14.71	7.000	5.440	4.626	3.961	3.487	3.023	(2.85)

R=138.0.

Our viscosity data are higher than those of Martin ⁽²²⁾ and of Bingham ⁽²³⁾ over comparable ranges (0 to 100°).

Phenyl acetonitrile.—B. pt. 231.0° C. at 755.8 mm. $A = -0.7593$;
B=605.2, from 90° upwards.

Temp.	18.2	90.2	120.5	140.3	160.7	180.5	200.1	231.0
D	1.0149	0.9567	0.9321	0.9153	0.8971	0.8796	0.8622	(0.8347)
1000 η	23.25	8.046	6.005	5.083	4.303	3.754	3.312	(2.76)

R=158.3.

Di-ethylamine.—B. pt. 55.85 to 55.95° C. at 747.8 mm. $A = -1.0777$;
B=470.6, over the range studied.

Temp.	10.2	37.6	44.4	51.7	55.95
D	0.7195	0.6909	0.6827	0.6756	(0.6710)
1000 η	3.878	2.732	2.540	2.357	(2.25)

R=119.6 ; calculated 119.8.

Our interpolated value for the viscosity at 35°, namely 2.82, is close to that of Kurnakov ⁽²⁴⁾ (2.79).

Di-n-propylamine.—B. pt. 108.9° C. at 752.9 mm. The plot of $\log 10^3\eta$ against $1/T$ is not linear.

Temp.	20.1	37.2	52.3	64.5	79.5	93.3	103.5	108.9
D	0.7374	0.7218	0.7083	0.6969	0.6830	0.6700	0.6607	(0.6556)
1000 η	5.335	4.267	3.610	3.179	2.744	2.422	2.215	(2.116)

R=167.8 ; calculated 167.4.

Di-n-butylamine.—B. pt. 159.0° C. at 748.7 mm. $A = -0.9534$; $B = 553.2$, from 100° upwards.

Temp.	58.9	83.1	101.3	110.6	121.0
D	0.7287	0.7083	0.6931	0.6854	0.6768
1000 η	5.343	3.983	3.354	3.090	2.835
Temp.	130.3	139.7	151.2	159.0	
D	0.6686	0.6606	0.6508	(0.6441)	
1000 η	2.626	2.441	2.242	(2.124)	

$R = 218.0$; calculated 215.0.

Methyl aniline.—B. pt. 195.5 to 196.0° C. at 753.6 mm. $A = -0.9200$; $B = 640.0$, over the range studied.

Temp.	95.5	122.8	150.0	180.5	196.0
D	0.9261	0.9035	0.8800	0.8535	(0.8400)
1000 η	6.585	4.950	3.913	3.108	(2.78)

$R = 144.0$; calculated 145.6.

Our temperatures do not overlap those of Bingham⁽²³⁾ or of Thole⁽²⁵⁾.

Ethyl aniline.—B. pt. 204.7 to 205.0° C. at 739.2 mm. $A = -1.0033$; $B = 664.3$, from 130° upwards.

Temp.	99.5	109.5	121.4	134.4	146.1	157.3
D	0.8939	0.8856	0.8751	0.8637	0.8530	0.8421
1000 η	5.981	5.401	4.792	4.240	3.816	3.466
Temp.		170.2	185.0	196.3	205.0	
D		0.8295	0.8149	0.8037	(0.7950)	
1000 η		3.127	2.801	2.585	(2.436)	

$R = 169.0$; calculated 169.4.

Bingham's⁽²³⁾ value (6.037) for the viscosity at 100° (his highest temperature) compares with our interpolated value of 5.96.

n-Propyl aniline.—B. pt. 222° C. at 765 mm. $A = -0.9585$; $B = 669.6$, from 130° upwards.

Temp.	106.5	136.0	150.9	165.7	179.8	194.5	210.5	222.0
D	0.8767	0.8520	0.8398	0.8272	0.8152	0.8023	0.7876	(0.7771)
1000 η	6.458	4.778	4.183	3.698	3.320	2.978	2.667	(2.48)

$R = 193.2$; calculated 193.2.

n-Butyl aniline.—B. pt. 239.5° C. at 749.7 mm. $A = -0.9576$; $B = 683.3$, from 140° upwards.

Temp.	98.9	130.9	144.2	160.3	174.6
D	0.8705	0.8460	0.8357	0.8228	0.8111
1000 η	7.60	5.440	4.783	4.163	3.712
Temp.	193.1	207.2	220.6	239.5	
D	0.7959	0.7837	0.7721	(0.7556)	
1000 η	3.226	2.917	2.673	(2.375)	

$R = 218.0$; calculated 217.0.

Piperidine.—B. pt. 103.5 to 105.2° C. at 747.5 mm. $A = -1.2620$; $B = 722.8$, from 60° upwards.

Temp.	13.55	51.1	64.5	76.6	86.1	95.2	105.2
D	0.8826	0.8479	0.8353	0.8239	0.8146	0.8055	(0.7955)
1000 η	20.75	9.340	7.580	6.377	5.600	5.025	(4.457)

$R = 128.2$; calculated 127.0, assuming the rheochor of the ring is the same as that of a saturated 6-carbon ring, namely -5.6 .

Tri-ethylamine.—B. pt. 85.9°C . at 755.0 mm. $A = -0.8224$; $B = 439.2$, from 60° upwards.

Temp.	12.52	49.6	60.7	70.5	80.0	85.9
D	0.7348	0.6999	0.6900	0.6794	0.6704	(0.6650)
1000 η	3.704	2.760	2.520	2.316	2.137	(2.028)

$R = 166.0$; calculated 165.9.

Di-methyl aniline.—B. pt. 193.0°C . at 754.9 mm. $A = -0.8490$; $B = 589.3$, from 150° upwards.

Temp.	86.9	121.9	135.0	150.8	166.2	180.1	189.2	193.0
D	0.9013	0.8726	0.8614	0.8482	0.8344	0.8216	0.8130	(0.8094)
1000 η	5.953	4.354	3.923	3.477	3.111	2.824	2.667	(2.603)

$R = 167.4$; calculated 167.9.

Our interpolated viscosities are lower than those of Bramley⁽²⁶⁾ but virtually identical with those of Bingham⁽²³⁾ at comparable temperatures (80 to 100°).

Methyl ethyl aniline.—B. pt. 204.7°C . at 742.2 mm. $A = -0.9400$; $B = 633.93$, from 130° upwards.

Temp.	109.3	122.4	134.8	147.5	160.0
D	0.8693	0.8587	0.8483	0.8372	0.8257
1000 η	5.142	4.593	4.119	3.696	3.343
Temp.	172.0	185.7	195.0	204.7	
D	0.8142	0.8008	0.7916	(0.7821)	
1000 η	3.054	2.764	2.595	(2.437)	

$R = 191.5$; calculated 191.7.

Di-ethyl aniline.—B. pt. 215.5 to 216.3°C . at 754.7 mm. $A = -0.7686$; $B = 583.0$, from 100° upwards.

Temp.	15.9	100.3	140.0	162.0	188.0	216.3
D	0.9383	0.8711	0.8412	0.8225	0.8032	(0.7813)
1000 η	24.74	6.210	4.394	3.676	3.132	(2.65)

$R = 213.7$; calculated 215.5.

Our calculated value for the viscosity at 98° is 6.35, and compares favourably with Bingham's⁽²³⁾ 6.31 (his highest temperature).

Di-methyl p-toluidine.—B. pt. 210.5°C . at 745.6 mm. $A = -0.8262$; $B = 578.1$, from 130° upwards.

Temp.	101.0	114.8	130.9	145.1	160.5
D	0.8728	0.8618	0.8486	0.8366	0.8232
1000 η	5.215	4.608	4.027	3.601	3.217
Temp.	175.0	190.0	204.2	210.5	
D	0.8112	0.7987	0.7867	(0.7814)	
1000 η	2.910	2.646	2.423	(2.341)	

$R = 190.8$; calculated 191.7.

Di-methyl o-toluidine.—B. pt. 185.0°C . at 749.5 mm. $A = -0.9422$; $B = 621.23$, above 150° .

Temp.	72.3	92.8	110.0	126.6	142.2
D	0.8845	0.8672	0.8519	0.8385	0.8247
1000 η	7.050	5.580	4.689	4.073	3.574
Temp.	155.0	170.0	181.5	185.0	
D	0.8132	0.7996	0.7890	(0.7858)	
1000 η	3.231	2.885	2.659	(2.596)	

$R = 192.1$; calculated 191.7.

Di-ethyl m-toluidine.—B. pt. 230.0° C. at 754.9 mm. $A = 0.9441$;
 $B = 666.7$ above 150°.

Temp.	114.7	135.0	155.5	165.4	176.0
D	0.8501	0.8332	0.8160	0.8076	0.7984
1000 η	5.947	4.904	4.095	3.773	3.467
Temp.	185.3	194.7	206.2	218	230.0
D	0.7904	0.7825	0.7722	0.7620	(0.7516)
1000 η	3.237	3.030	2.796	2.591	(2.406)

$R = 239.9$; calculated 239.3.

References.

- (1) Friend and Hargreaves, *Phil. Mag.* xxxv. p. 57 (1944).
- (2) Friend and Hargreaves, *Phil. Mag.* xxxiv. p. 643 (1943).
- (3) Glasstone, 'Recent Advances in Physical Chemistry' (Churchill, 1931), p. 67.
- (4) Sugden, 'The Parachor and Valency' (Routledge, 1930), p. 45. Mumford and Phillips (*J. Chem. Soc.* p. 2112 (1929)), in their interpretation of the parachor do distinguish between H in CH and OH. Their figures would support formula I.
- (5) For "acetone" see Bramley, *J. Chem. Soc.* cix. p. 11 (1916). For "methyl ethyl ketone" see ref. (2) above. For the other ketones see ref. (9) below.
- (6) Fredenhagen and Dahlmlos, *Zeitsch. anorg. Chem.* clxxxix. p. 77 (1929), quoted in 'International Critical Tables,' vii. p. 213.
- (7) G. E. Coates and J. E. Coates, *J. Chem. Soc.* p. 77 (1944).
- (8) Friend and Hargreaves, *Phil. Mag.* xxxv. p. 136 (1944).
- (9) Thorpe and Rodger, *Phil. Trans.* clxxxv. p. 570 (1895).
- (10) Elsey, *J. Amer. Chem. Soc.* xlii. p. 2454 (1920). Fitzgerald, *J. Physical Chem.* xvi. p. 621 (1912).
- (11) Plank and Hunt, *J. Amer. Chem. Soc.* lxi. p. 3590 (1939).
- (12) Auger, *Compt. rend.* cxlv. p. 1289 (1907).
- (13) Adams and Marvel, *J. Amer. Chem. Soc.* xlii. p. 310 (1920).
- (14) Maass and Boomer, *ibid.* xlv. p. 1709 (1922).
- (15) Muchin, *Zeitsch. Elektrochem.* xix. p. 809 (1913), quoted by I. C. T. vii. p. 218.
- (16) Efremov, *J. Russ. Phys. Chem. Soc.* l. p. 340 (1918), quoted by I. C. T. vii. p. 213.
- (17) Creighton, *J. Franklin Inst.* clxxxix. p. 641 (1920), quoted by I. C. T. vii. p. 215.
- (18) Walden, *Zeitsch. physikal. Chem.* lv. p. 207 (1906).
- (19) Walden and Birr, *ibid.* cxliv. p. 269 (1929).
- (20) Muchin and Faermann, quoted (without reference) in Landolt, *Physikalisch-Chemische Tabellen*, Eg. II. p. 98 (1931).
- (21) Sachanov and Rabinovich, *J. Russ. Phys. Chem. Soc.* xlvii. p. 81 (1915).
- (22) Martin, *J. Chem. Soc.* p. 3270 (1928).
- (23) Bingham, van Klooster and Kleinspehn, *J. Physical Chem.* xxiv. p. 1 (1920).
- (24) Kurnakov and Zhemchuzhnyi, *Zeitsch. physikal. Chem.* lxxxiii. p. 481 (1913).
- (25) Thole, *J. Chem. Soc.* ciii. p. 320 (1913).
- (26) Bramley, *ibid.* cix. p. 10 (1916).

The Technical College,
 Birmingham.

LXXII. *On the Optimum Thickness of Powder Specimens in X-ray Diffraction Work.*

By A. TAYLOR, M.Sc., Ph.D., F.Inst.P.†

[Received November 10, 1943.]

Introduction.

OF the many factors which influence the intensities of the spectra from powder specimens used in X-ray diffraction work, the absorption of the incident and reflected beams within the specimen plays a very important part. Before and after reflection, "macro"-absorption takes place along the incident and emergent paths within the main bulk of the specimen. These paths have a mean linear absorption coefficient which takes into account the voids between the particles of powder. Account must also be taken of the absorption taking place within the individual particles which reflect the beam. This is termed the microabsorption. The smaller the polycrystalline particles composing the specimen, the greater is the energy which can be reflected into the Debye-Scherrer spectra ⁽¹⁾.

With particles of given size, the intensity of the reflections depends on the total volume of irradiated powder. In the case of flat powder compacts, the incident beam and the reflected transmission spectra pass right through the thickness of the specimen. Thus the intensities of the spectra are proportional to the thickness of the compact. The spectra are, at the same time, reduced in intensity by a function of $e^{-\mu z}$, where μ is the average linear absorption coefficient of the powder compact and z the total length of path of a reflected ray before and after reflection. Starting from a vanishingly thin specimen, the intensity of a transmitted reflection will increase with increase in thickness until the exponential factor becomes important enough to produce a diminution in the transmitted energy. There is thus an optimum thickness for any powder specimen which produces transmission Debye-Scherrer patterns.

In the case of surface-reflection and back-reflection diagrams from flat powder compacts, increasing the thickness results in a continued increase of the spectral intensities up to a limiting value set by the sensitivity of the film or recording device. At this limit the plate is considered to behave as if it were infinitely thick.

The above cases are well known and easy to calculate. The same methods may also be applied to determine the optimum radius for a cylindrical Debye-Scherrer specimen. In this instance it is not possible

† Communicated by the Author.

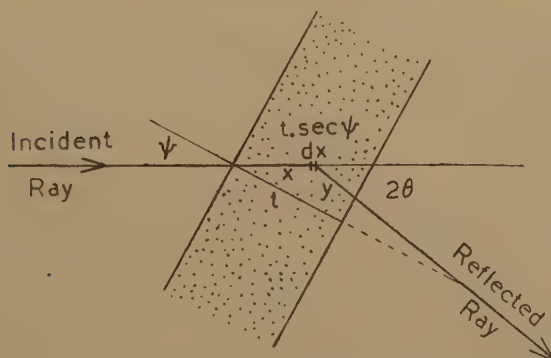
to write down an expression from which the best radius can be obtained by direct computation and a graphical method must be resorted to.

1. The Optimum Thickness of a Flat Powder Compact for Transmission Spectra.

We shall consider the general case of a flat powder compact of mean linear absorption coefficient μ and thickness t , irradiated by a parallel beam of monochromatic X-rays which make an angle ψ with the normal to the plate (fig. 1). Since the cross-section of the incident beam is fixed by the slit system of the camera, any volume dV of the specimen swept out by the incident beam will be proportional to the depth of penetration dx . We may therefore confine our attention to a single ray which is deviated through an angle 2θ at various points in its path by the crystallites in the specimen, θ being the Bragg angle.

The transmitted intensity dI is proportional to the reflecting length dx and to $e^{-\mu(x+y)}$. Summing for all particles along the length $t \sec \psi$

Fig. 1.



Paths of incident and transmitted rays through flat powder compact.

swept out by the incident ray, the intensity of the transmitted reflection becomes

$$I = \text{const. } I_0 \int_0^{t \sec \psi} e^{-\mu(x+y)} dx. \quad (1)$$

Now from the geometry of the figure

$$y = t \sec (2\theta - \psi) - x \cos \psi \sec (2\theta - \psi).$$

Writing

$$a = \sec (2\theta - \psi), \quad b = \sec \psi, \quad c = \frac{1}{b} = \cos \psi,$$

we obtain

$$\begin{aligned} I &= \text{const. } I_0 \int_0^{bt} e^{-\mu(x-ax+at)} dx, \\ &= \frac{\text{const. } I_0}{\mu(ac-1)} \{e^{-\mu bt} - e^{-\mu at}\}. \end{aligned}$$

The optimum thickness t^* occurs when $dI/dt=0$, i. e. when

$$\mu t^* = \frac{1}{a-b} \log_e \frac{a}{b} \quad \dots \dots \dots (2)$$

or
$$\mu t^* = \frac{1}{\sec(2\theta-\psi) - \sec\psi} \cdot \log_e \left\{ \frac{\sec(2\theta-\psi)}{\sec\psi} \right\} \quad \dots \dots \dots (3)$$

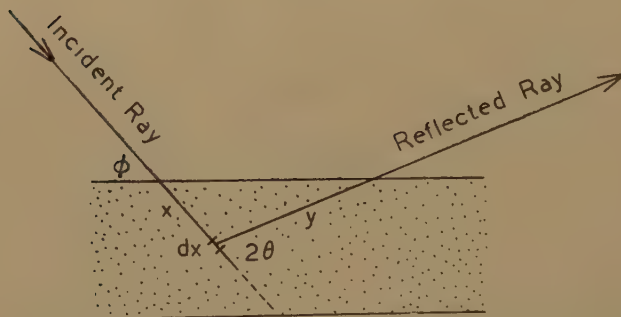
Equation (3) gives the optimum value of t for the most general case. For special cases it is convenient to write equation (3) in the form

$$\begin{aligned} \mu t^* &= \frac{1}{\sec(2\theta-\psi) - \sec\psi} \log_e \left\{ 1 + \frac{\sec(2\theta-\psi) - \sec\psi}{\sec\psi} \right\}, \\ &= \frac{1}{\sec\psi} - \frac{1}{2} \frac{\sec(2\theta-\psi) - \sec\psi}{\sec^2\psi} + \dots \dots \dots (4) \end{aligned}$$

(a) *Special Case of Normal Incidence.* When the incident beam is normal to the face of the specimen, $\psi=0^\circ$ and equation (3) takes the form

$$\mu t^* = \frac{1}{\sec 2\theta - 1} \cdot \log_e \sec 2\theta \quad \dots \dots \dots (5)$$

Fig. 2.



Paths of incident and reflected rays in surface reflection arrangement.

(b) *Special Case when $\psi=\theta$.* In this case the plane of the specimen bisects the angle $180^\circ-2\theta$ between the incident and transmitted rays. Using equation (4) and putting $\psi=\theta$, we immediately obtain

$$\mu t^* = \frac{1}{\sec \theta} \quad \dots \dots \dots (6)$$

(c) *Special Case when $\theta \rightarrow 0^\circ$.* For very small Bragg angles, and $\psi=0^\circ$, we obtain

$$\mu t^* = 1 \quad \dots \dots \dots (7)$$

II. Optimum Thickness of Flat Powder Plate for Reflection Spectra.

In fig. 2, let the incident ray strike the plate specimen at the grazing angle ϕ and be deviated through the angle 2θ after reflection,

The total intensity I in the reflected ray is

$$I = \text{const. } I_0 \int_0^x e^{-\mu(x+y)} dx,$$

and since

$$y = \frac{x \sin \phi}{\sin (2\theta - \phi)},$$

$$\begin{aligned} I &= \text{const. } I_0 \int_0^x e^{-\mu x \left(1 + \frac{\sin \phi}{\sin (2\theta - \phi)}\right)} dx, \\ &= \frac{\text{const. } I_0}{\mu \left(1 + \frac{\sin \phi}{\sin (2\theta - \phi)}\right)} \left\{ 1 - e^{-\mu x \left(1 + \frac{\sin \phi}{\sin (2\theta - \phi)}\right)} \right\}. \quad (8) \end{aligned}$$

Thus I increases as x increases reaching a limiting value when x is effectively "infinitely thick," in which case

$$I = \frac{\text{const. } I_0}{\mu} \cdot \frac{\sin (2\theta - \phi)}{[\sin (2\theta - \phi) + \sin \phi]}. \quad (9)$$

The criterion of infinite thickness is set by the sensitivity of the X-ray film or recording device. For suppose the crystals at the surface give rise to reflections of intensity $I(s)$ and those at a depth x give an intensity $I(x)$. The reflections of $I(x)$ will not be visible if $I(x)$ is less than $\frac{I(s)}{600}$ approximately. This ratio is taken on the same basis as when the strongest $K\beta$ reflection is invisible after filtration has cut down the $K\beta$ component to $\frac{1}{600}$ of $K\alpha$ (2). This may be considered by some to be over-generous.

We thus have

$$\begin{aligned} \frac{I(s)}{I(x)} &= \frac{600}{1} = \frac{\text{const. } I_0 e^{-\mu \cdot 0}}{\text{const. } I_0 e^{-\mu(x+y)}} \\ &= e^{\mu x \left(1 + \frac{\sin \phi}{\sin (2\theta - \phi)}\right)}. \end{aligned}$$

Hence

$$\mu x = \frac{6.4 \sin (2\theta - \phi)}{[\sin (2\theta - \phi) + \sin \phi]}. \quad (10)$$

In back-reflection patterns, $\phi = 90^\circ$ and $\theta \rightarrow 90^\circ$, in which case the condition for infinite thickness is that x should exceed the value given by

$$\mu x = 3.2 \quad (11)$$

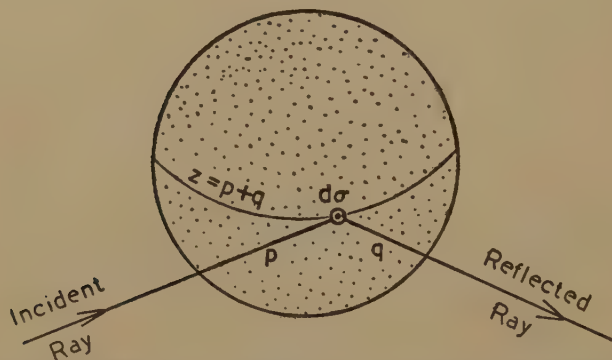
III. Optimum Thickness of a Cylindrical Powder Specimen.

Consider the cylindrical Debye-Scherrer powder specimen shown in cross-section in fig. 3. The intensity reflected from the element of volume dV will be cut down by the amount $e^{-\mu z}$ where $z = p + q$. The

total intensity I in the reflected beam from the entire specimen will therefore be

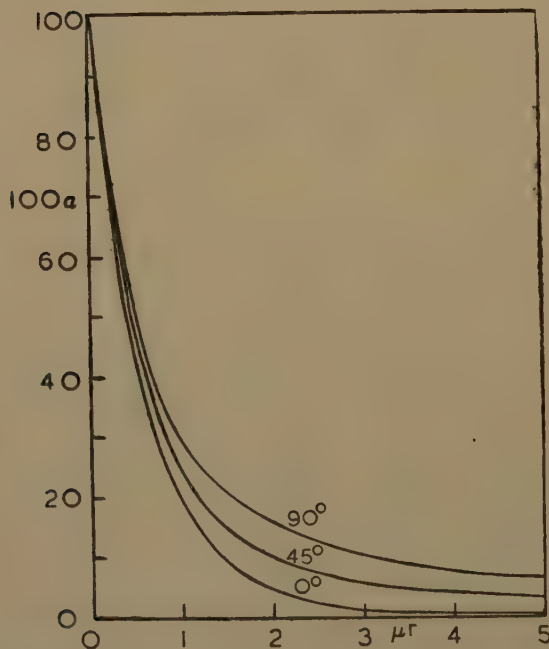
$$I = \text{const. } I_0 \iiint e^{-\mu z} dV. \quad \dots \quad (12)$$

Fig. 3.



Paths of incident and reflected rays through cylindrical powder specimen.

Fig. 4.



Values of $100\alpha = 100 \int e^{-\mu r x} d\sigma$ for cylindrical Debye-Scherrer powder specimens as a function of μr and Bragg angle.

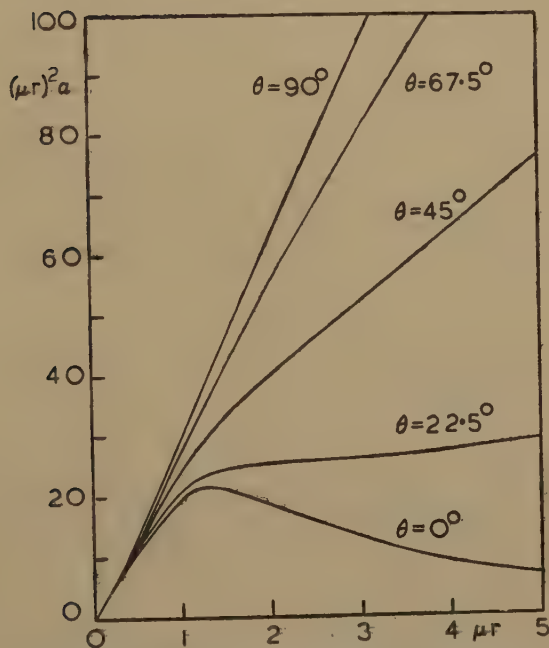
Replacing dV by $h d\sigma$, where h is the height of the cylinder and $d\sigma$ an element of cross-sectional area, equation (12) becomes

$$I = \text{const. } I_0 h \iint e^{-\mu z} d\sigma.$$

Further, if r is the radius of the specimen, we may write

$$\begin{aligned} I &= \text{const. } I_0 h \cdot \pi r^2 \iint e^{-\mu r x} ds, \\ &= \text{const. } I_0 h \frac{\pi}{\mu^2} (\mu r)^2 \iint e^{-\mu r x} ds, \end{aligned}$$

Fig. 5.



Curves of $(\mu r)^2 \alpha$ for reflections from cylindrical powder specimens at various Bragg angles.

$$= \text{const. } I_0 \frac{h\pi}{\mu^2} (\mu r)^2 \alpha, \dots \dots \dots (13)$$

where $ds = \frac{d\sigma}{\pi r^2}$, and $x = z/r$.

I will attain a maximum value when $(\mu r)^2 \iint e^{-\mu r x} ds$ is a maximum. The value of μr at which this occurs cannot be found by simple differentiation, as in the previous cases of the flat compacts, since $\iint e^{-\mu r x} ds$ is not a simple function of μr . To find the value of the integral we have to use the method of Claassen⁽³⁾ and divide the cross-section into loci

such that z is a constant. This has been carried out analytically by Bradley⁽⁴⁾ for a large range of μr values for special Bragg angles. Intermediate values are obtained by interpolation. Values of $\alpha = \iint e^{-\mu r x} ds$ are shown in fig. 4.

From these curves of α we can construct curves of $(\mu r)^2 \alpha$ as shown in fig. 5 for $\theta = 0^\circ, 22\frac{1}{2}^\circ, 45^\circ, 67\frac{1}{2}^\circ$ and 90° .

At 0° there is a well-marked maximum in the region of $\mu r = 1.33$. At $\theta \approx 20^\circ$ the function becomes almost stationary beyond $\mu r = 1.4$. In this region any increase in specimen radius or concentration of the powder produces no increases in X-ray intensity. For angles above 20° , the efficiency of reflection of the specimen continues to increase with increasing μr , partly on account of the greater thickness of the specimen and partly on account of the greater surface area exposed to the incident radiation, a condition not met with in the case of a flat compact irradiated by a beam of limited cross-section. A metal specimen in wire form yields very sharp, weak reflections at low angles and relatively strong reflections at high angles. A specimen of the same material in powder form and suitably diluted with Canada Balsam yields spectra of the same order of intensity over the whole angular range, but in this case the spectra in the lower orders are less sharp than those given by the wire.

Acknowledgment.

The author wishes to thank the English Electric Co. Ltd. for permission to publish the above article.

References.

- (1) A. Taylor, Phil Mag. xxxv. p. 215 (1944). Also, correspondence by G. W. Brindley and A. Taylor, Phil. Mag. xxxv. pp. 638-641.
- (2) O. S. Edwards and H. Lipson, J. Scientific Instr. xviii. p. 131 (1941).
- (3) A. Claassen, Phil. Mag. (7), ix. p. 57 (1930).
- (4) A. J. Bradley, Proc. Phys. Soc. xlvii. p. 879 (1935).

LXXIII. The Effect of Heterogeneity in Powders on the Intensity of X-ray Reflections.

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

IN two recent papers in this Journal, Dr. A. Taylor (1944 *a, b*) discusses the effect of heterogeneity in crystalline powders on the intensities of X-ray reflections evidently without fully realizing the extent to which the subject has previously been considered, for example by Brentano (1935, 1938), and by Brindley and Spiers (1938).

Dr. Taylor's thesis is that *crystal* size must be taken into account in deriving an expression for the X-ray intensity reflected by a powder when there is appreciable absorption in an individual crystal. In his second paper, he discusses, *inter alia*, measurements by Brindley and

Spiers (1935) of reflected intensities from filed metal powders and attempts to show that their results can be explained in terms of a crystal absorption factor derived in the first paper.

The writer contends that differential absorption effects may arise when X-rays are reflected by *mixed* powders owing to the heterogeneity of the powder specimen produced by the finite *particle* size (not crystal size) of the constituents if the particles individually absorb an appreciable amount of radiation. With powders of *single* substances (such as Brindley and Spiers employed in their experiments on distortion in filed metal powders) the question of differential absorption does not arise. Briefly it may be stated that when X-rays are reflected by a mixed powder, the absorption coefficient which determines the reflected intensity from any particular crystal is characteristic of the mixture over the whole of the path of the radiation in the powder except that part which lies within the particle to which the reflecting crystal belongs. The correction factor involved depends on the size of the particles and on the difference between the absorption coefficients of the separate constituents and of the mixture as a whole. The argument will be given in more detail in a paper shortly to be communicated to the 'Philosophical Magazine.'

The method and numerical tables given by Dr. Taylor for calculating his crystal absorption factor for spherical crystals can equally well be used for calculating the particle absorption factor for spherical particles. The writer, therefore, does not question the usefulness of this section of Dr. Taylor's first paper. As regards the second paper, it is considered that the argument regarding Brindley and Spiers's measurements is wholly invalid because differential absorption effects cannot arise with single powders.

In his first paper Dr. Taylor has performed a useful service in stressing the important effect of grain size in alloys having two or more phases on the relative X-ray intensities reflected by the separate components. Although in this case no distinction exists between crystals and particles as regards the heterogeneity of the solid material, the form of the correction factor given by Dr. Taylor still requires modification as will be indicated in the writer's paper.

References.

- Brentano, J. C. M., Proc. Phys. Soc. xlvii. p. 932 (1935).
- Brentano, J. C. M., Proc. Phys. Soc. l. p. 247 (1938).
- Brindley, G. W., and Spiers, F. W., Phil. Mag. xx. p. 882 (1935).
- Brindley, G. W., and Spiers, F. W., Proc. Phys. Soc. l. p. 17 (1938).
- Taylor, A., Phil. Mag. xxxv. p. 215 (1944 a).
- Taylor, A., Phil. Mag. xxxv. p. 404 (1944 b).

Physics Laboratories,
University of Leeds.
August 28th, 1944.

Yours faithfully,
G. W. BRINDLEY.

LXXIV. *Reply to Dr. Brindley's letter on "The Effect of Heterogeneity in Powders on the Intensity of X-ray Reflections."*

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

THE points raised by Dr. Brindley in his letter are most interesting. His paper, which is in the course of preparation and which I have had the privilege of discussing with him, should do much to clarify this difficult problem.

First of all, there is the question of the correct value of the absorption coefficient which should be applied to the reflecting crystal or polycrystalline particle. After a survey of the literature, which was inaccessible at the time of developing my theory, I find that, in addition to myself, Schäfer⁽¹⁾, Brentano⁽²⁾ and even Brindley and Spiers^{(3), (4)} have used the linear absorption coefficient μ of the particle when computing the effect of particle size on the reflected intensity. In fact, Brindley and Spiers⁽³⁾ say "the agreement is sufficiently good to indicate clearly the magnitude of the effect of particle size on results obtained by the mixture method and to show that Schäfer's expressions give a fairly reliable estimate of the degree of accuracy to be expected with any given mixture." Rusterholz⁽⁵⁾ alone uses the difference between the linear absorption coefficient of the reflecting particle, μ , and the mean linear absorption coefficient of the (solid) surrounding powder mixture, $\bar{\mu}$.

The second point concerns the use of particle radius as distinct from crystal radius. My own discussion⁽⁶⁾ selects the reflecting crystal as the unit and includes the remainder of the non-reflecting polycrystalline particle with the non-reflecting surrounding medium. The more detailed analysis given by Dr. Brindley shows that, in summing the effects of the reflecting crystals in any layer of the specimen, we do, in fact, "synthesise" particles having the same radius and linear absorption coefficients as the polycrystalline particles of the specimen. Thus, we must compute the value $(\mu - \bar{\mu})a$, where a is the radius of the polycrystalline particle, in order to look up the correct value of the micro-absorption factor τ in the tables. On this Dr. Brindley and I are in complete agreement.

If the powder consists of a single substance, $(\mu - \bar{\mu})a = 0$ and the value of τ is unity. Thus, as Dr. Brindley points out in his letter, the micro-absorption factor cannot influence his results on filed single powders of copper and nickel, as contended in my second paper.⁽⁷⁾ This only becomes fully evident when Rusterholz's theory is applied. In its revised form, the micro-absorption factor still applies to the analysis of mixed powders as given in my first paper.

If there is appreciable secondary extinction, the effective absorption coefficient of the individual crystal in its reflecting position is somewhat

greater than that of the particle surrounding it, having now a value $\mu + gQ$ where gQ is a function of the intensity of the reflected beam. It would thus appear that to allow for extinction the absorption factor A for the specimen should be written in the form

$$A = \alpha \tau \epsilon,$$

where α is the macro-absorption factor, τ the micro-absorption factor for the polycrystalline particle given in the table in terms of $(\mu - \bar{\mu})a$, and ϵ is the extinction factor for the individual reflecting crystallite of radius a' . ϵ is obtained from the τ -tables in terms of $(\mu + gQ - \bar{\mu})a'$. With a single substance, $\mu - \bar{\mu} = 0$ and the absorption factor becomes simply $A = \alpha \epsilon$.

On the basis of the original argument, it appeared that the ratio of intensities of superlattice lines to main lattice lines of Cu_3Au was a function of wave-length and domain size.⁽⁷⁾ Since $\mu - \bar{\mu} = 0$, $\tau = 1$ and there cannot be any differential absorption effects. There still remains the residual effect due to secondary extinction, so that even with this modified theory we should still expect to find the superlattice lines to appear too strong in relation to the main spectra.

References.

- (1) K. Schäfer, *Zeits. für Physik*, xlv. p. 738 (1933).
- (2) J. C. M. Brentano, *Proc. Phys. Soc.* xlvii. p. 932 (1935).
- (3) G. W. Brindley and F. W. Spiers, *Proc. Phys. Soc.* l. p. 17 (1938).
- (4) G. W. Brindley and F. W. Spiers, *Phil. Mag.* xx. p. 882 (1935).
- (5) A. Rusterholz, *Helv. Phys. Acta*, iv. p. 68 (1931).
- (6) A. Taylor, *Phil. Mag.* xxxv. p. 215 (1944).
- (7) A. Taylor, *Phil. Mag.* xxxv. p. 404 (1944).

Yours faithfully,

A. TAYLOR,
English Electric Co., Ltd.

LXXV. Notice respecting New Book.

Colorimetric Methods for the Determination of Traces of Metals. By E. B. SANDELL. [Pp. xvi+486. 73 diagrams.] (Interscience Publishers, Inc. 1944). Price \$7.00 net.

DR. SANDELL has produced a most valuable addition to the series of monographs being published by the Interscience Publishers Inc. on Chemical Analysis. The essential principle in colorimetric methods is the comparison of the sample solution with a standard solution in respect of its colour or its spectral absorption. In instruments of the type of the Duboscq colorimeter, Beer's law may be assumed and the concentration derived by comparing the lengths l_s and l_p of the sample and standard solutions respectively which are found

to yield the same colour. The colour of the sample is in fact duplicated by the appropriate length of the standard solution, so that the term "colorimetry" in its strict sense of the measurement of colour is, as the author points out, a misnomer; its use is now so common, however, that the opportunity for its substitution by a better term has probably passed. In spectrophotometry, a monochromator or a colour filter transmitting a narrow band of wave-lengths is used to isolate a particular part of the spectrum, and the absorption of the sample is measured by a suitable photometric device at one or more selected wave-lengths. The absorption is then compared with that previously recorded for a series of standard concentrations and, by interpolation, the concentration of the sample can be deduced.

The successful application of these methods to the accurate determination of traces of metals depends (a) on the possibility of preparing solutions whose colour is highly sensitive to the concentration of the metal under test, (b) on the elimination of other metals and constituents which might affect the colour of the solution, and (c) on making observations of the colour with adequate precision and accuracy.

About four-fifths of Dr. Sandell's book is rightly devoted to detailed descriptions of the methods of isolation, preparation of solutions and the colorimetric determination appropriate to each metal and to the medium in which the metal is found, *e. g.*, in silicate rocks or biological material. This section is well documented and provides a valuable work of reference. The shorter section of the book deals with colorimetric methods in general, but it is of very special interest for the careful examination of the factors affecting the accuracy of measurement. The author does not omit to distinguish between precision and accuracy and endeavours, for example, to relate the sensitivity of the eye as expressed by Weber's law to the optimum conditions of observation in visual instruments. Both visual and photo-electric methods are considered and short accounts of fluorimetry and nephelometry are included. The remarkable accuracy which can be achieved in trace analysis by colorimetric methods is in many cases so high that it is equal to, or even surpasses, that of any other applicable method. And this in spite of the fact that the determination can often be made with inexpensive apparatus and with little preliminary training on the part of the analyst.

W. D. WRIGHT.

[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]

LXXVI. *A New Photo-Electric Method for Fourier Synthesis and Analysis.*

By R. FÜRTH and R. W. PRINGLE, Edinburgh *.

[Received September 28, 1944.]

Summary.

A model instrument, working on a new photo-electric principle, is described by which the synthesis of a function from its given Fourier components and the analysis of a given function into such components can be performed automatically. Used as a synthesiser the instrument produces the graph of the synthesised function instantaneously on the screen of a cathode ray oscillograph when the Fourier coefficients have been set, and the effect of a change in the values of the amplitudes and phases of the components on the shape of the graph can be readily demonstrated. Used as an analyser the instrument makes it possible to obtain the values of the Fourier coefficients from readings of the deflections of a microammeter when a disk bearing the shape of the given function in polar co-ordinates has been inserted. The instrument can be easily adapted to synthesis and analysis with respect to any other set of orthogonal functions.

§ 1.

THE great importance of both Fourier Analysis and Synthesis in pure and applied physics, *e. g.*, in crystallography, acoustics, mechanical and electrical engineering, geophysics and astrophysics has made it necessary to develop methods by which the synthesis of periodic functions from harmonic components and the analysis of given periodic functions into such components can be performed rapidly and, if possible, automatically. Consequently a large variety of excellent methods of this kind has been invented, some of which use tables or calculating and sorting machines ⁽¹⁾, and others mechanical, electrical and optical principles. A very good review on these methods is given in two articles by K. Bourne ⁽²⁾ and by L. J. Comrie ⁽³⁾. In the present paper a further method is proposed and shown to work satisfactorily, not because of any deficiency in the already existing methods, but because it is based on principles which have apparently not been used for this purpose before and therefore seems to be of some interest in itself, and also because one single instrument constructed on these principles can be used for both analysis and synthesis.

* Communicated by the Authors.

Owing to war conditions and other circumstances, the authors have only been able to build a model of such an instrument which has a number of deficiencies and hence by no means reaches the high accuracy and range of components of some other methods. But even in its present state it is quite useful for demonstration purposes, and there is no doubt that a proper technical development will bring its accuracy and range up to the level required for precision computations. Furthermore, the same instrument can by a slight modification (replacing a disk carrying a pattern of the component functions by another one) be used for analysis and synthesis with respect to any set of orthogonal functions ("generalized" Fourier Analysis and Synthesis), *e. g.*, Legendre functions, as the principle on which the method is based is by no means restricted to harmonic functions as those of most of the existing methods.

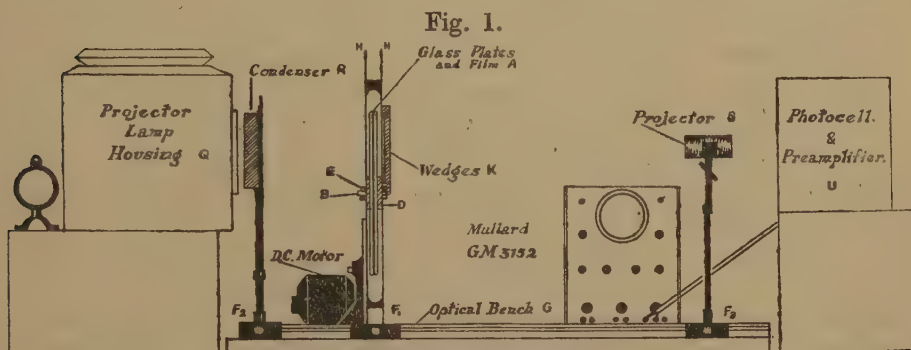
In the following the methods of synthesis and analysis are described separately and some experimental evidence for their proper functioning is given.

§ 2.

The principle of the Synthesiser consists in the production of a number of light beams the intensity of each of which is made to vary in time according to either a sine or cosine function with a frequency equal to an integer multiple of a fundamental frequency. By optical filters the amplitudes of these oscillations can be set so as to correspond to the given coefficients of a Fourier series. All the beams are concentrated on the light-sensitive cathode of a photo-electric cell; the time variation of the electric output of this cell therefore gives the synthesised function. It is amplified and applied to one pair of deflecting plates of an ordinary cathode ray oscillograph. A "linear time base" is supplied to the other pair of deflecting plates which is automatically synchronized to the fundamental frequency. Thus, provided that this frequency is sufficiently high, the trace of the spot on the screen of the cathode ray oscillograph will show immediately a graph of the synthesised function which can be observed directly for demonstration purposes, or measured on the screen, or finally traced or photographed for measuring purposes.

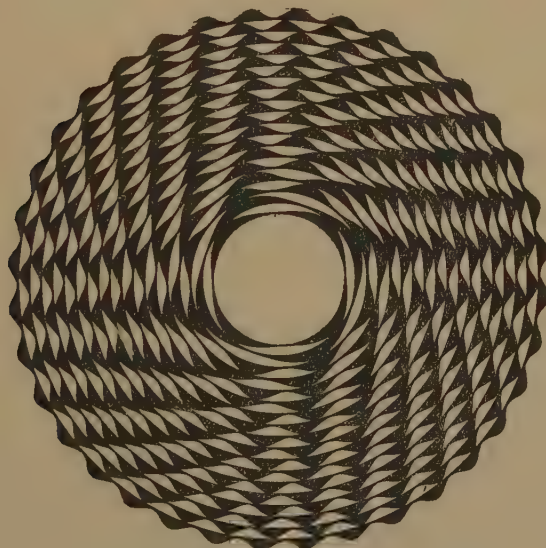
The whole apparatus is sketched in fig. 1. Its main part, used for modulating the light intensity, is the rotating disk A, a reduced photograph of which is shown in fig. 2. It is made of a photographic film, has a diameter of 21 cm., and contains 18 rings of 3.5 mm. width with interspaces of 0.75 mm. Each of these rings contains a variable area wave representing a harmonic function in polar co-ordinates, the two innermost being cosine and sine functions with four complete cycles, one in each quadrant. This gives the fundamental frequency of the series. The following rings are cosine and sine functions alternatively, whose frequencies, increasing with increasing radius, are integer multiples of the fundamental. A circular central region of the disk is cut out and is used for the analyser only (see § 3). This disk was produced by photo-

graphic reduction* (on slow "Kodakline" fine grain film) from an original drawn in Indian ink in large scale (112 cm. diameter). The film disk is mounted between two $\frac{1}{8}$ -inch thick circular glass plates with central holes. One of these plates is rigidly fixed to an axle B and carries on its circumference a cardboard ring C (fig. 3), inscribed with a scale, which at the same time serves as a frame for the film disk in which



The harmonic synthesiser.

Fig. 2.



Disk with patterns of harmonic functions.

the latter fits accurately. This ensures its precise centering with respect to the axle and is obviously most important. The other glass plate is used to hold the film in position and is itself secured with a lock-nut D (fig. 1).

* The photographic reproduction was very successfully performed by Messrs, Bartholomew, Edinburgh.

The axle of the disk turns in bearings E mounted on a clamp F_1 which is fixed to an optical bench G. The axle also carries a small wheel (not shown in the figures) which is driven by means of a belt by a 1/16 H.P. electromotor over a suitable reduction gear. The speed of the motor is regulated by a variable series resistance so that the disk normally rotates at about 300 r.p.m. The clamp F_1 also carries two vertical metal plates H serving as light screens between which the disk rotates. One of these plates (the one facing the light source) carries a vertical adjustable slit J of 10 cm. length and 0.5 mm. width which is adjusted so as to coincide exactly with the upper vertical radius of the disk in order to illuminate a narrow region of all the rings along this radius. The other of the plates H carries the system of filters for reducing the intensity of this illumination in each of the rings separately, corresponding to the given Fourier coefficients.

These filters consist of 18 neutral optical wedges K produced photographically on film strips (slow "Kodakline" fine grain film) by a simple method, and mounted between strips of Perspex * L which, apart from protecting the wedges, serve as runners for continuously moving the wedges across the slit. These runners are supported by a system of metal holders M, N in such a way that they can be easily moved independently of each other. The position of each wedge can be measured by reading the position of the edge of the corresponding runner on a scale O (visible on the left-hand side of fig. 3). Thus, provided the wedges have been calibrated, the coefficients of the Fourier series to be synthesised can be set by means of this scale. However, it will be shown later that a much more convenient method can be used which does not require any calibration of the filters.

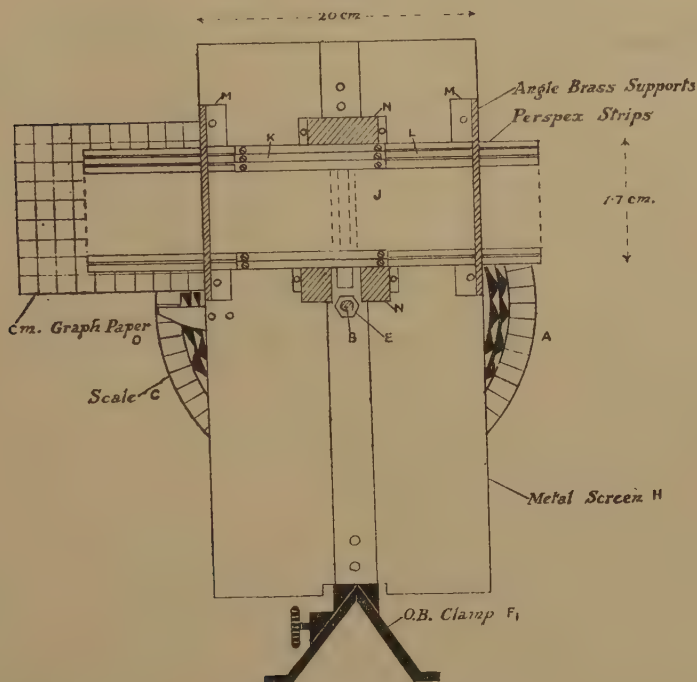
The light source consists of a 200 watt (Mazda) projector lamp P (fig. 4) mounted within a lantern Q (shown in fig. 1). A large condenser lens system (11 cm. diameter) R is placed between the lamp and the slit J, carrying a fine ground glass screen which serves as a diffuser in order to make the illumination of the slit as uniform as possible over its whole length. The slit and system of rotating rings thus form eighteen light sources the intensities of which vary harmonically in time according to the nine sine and the nine cosine functions. Reduced real images of these sources are focused on the photocell T by a projector lens system S with a reduction factor of about $\frac{1}{3}$. Both lenses R and S are mounted on the optical bench G by means of the clamps F_2 and F_3 .

A commercial Osram caesium type gasfilled photoelectric cell with an operating voltage of about 100 volts is used. Such a cell has the disadvantage that in order to obtain the highest sensitivity the voltage applied must be near the critical value at which glow discharge starts. This might cause non-linearity of the electric response of the cell and irregular fluctuations of the photo-current, usually described as "noise."

* The authors are indebted to I.C.I. (Plastics) for supplying the Perspex sheets and the necessary polishing material,

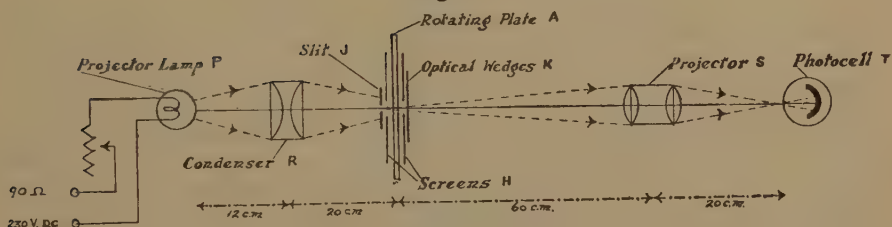
This could be avoided by using a secondary emission type vacuum photocell which unfortunately was not available. However, in the case of the synthesiser the sensitivity was sufficiently high to allow the cell to be operated at a voltage of only about 70 volts where none of the effects mentioned occurred.

Fig. 3.



Optical wedge assembly.

Fig. 4.



The optical system.

The anode current variations of the photocell have to be amplified considerably. This is done by inserting a resistance of 15 megohms in the anode circuit; the voltage across this resistance is applied to the control grid of a "pre-amplifier" consisting of an indirectly heated pentode valve (Osram VMP4G) which gives a voltage amplification of about 30 times. In order to suppress all disturbances due to stray

electric and magnetic fields the cell, the pre-amplifier, and the high tension batteries are placed in a metal box U which also shields the cell from indirect light; the light from the projector lamp S enters the box through a small window at the front. The filament accumulators for the valve are also placed in a metal box, and the leads between the two boxes shielded by metal tubes.

The cathode ray oscillograph used was a Mullard GM 3152 instrument which contains a two stage resistance-capacity coupled amplifier with push-pull output to the "Y-deflecting plates," the frequency response of which is sensibly flat down to 10 c/sec. The voltage across a resistance of 25,000 ohms in the pre-amplifier anode circuit is applied through screened leads to this amplifier the input impedance being about 1 megohm. The amplification provided is sufficient for the light from one of the harmonic functions to produce a vertical trace of about 3 cm. in length on the oscillograph screen.

The linear time base for the "X-deflecting plates" is provided by a D.C. source of 400 volts and a continuously rotating wire wound potentiometer of 10^5 ohms resistance attached to the axle B of the disk A. By this procedure obviously the time base becomes automatically synchronized with one-fourth of the fundamental frequency of the harmonic series, independently of the speed of the motor, thus producing a graph of the synthesised function covering four complete periods. By using a fraction of the voltage mentioned this graph can be adjusted so as just to cover the oscillograph screen; by using the whole voltage the figure is broadened and about $1\frac{1}{2}$ periods are produced on the screen.

Instead of using the calibration curves of the optical wedges and the scale O for setting the Fourier coefficients another procedure had been adopted, as already indicated. All the filters except that corresponding to the fundamental, say, sine wave are set to zero transparency, and the latter one adjusted so as to produce a wave of a convenient amplitude, say 3 cm., on the oscillograph screen. Then this beam is covered up without altering the position of the wedge, and the next component is adjusted so as to produce a wave with an amplitude in the correct proportion to that of the fundamental, by using a transparent square raster in front of the screen. This procedure is repeated until all the components are set according to the given Fourier coefficients. Finally, all the beams are uncovered and the correct shape of the synthesised function appears at once on the screen. By adopting this scheme all possible distortions due to lack of uniformity either of the frequency response of the amplifier or of the illumination of the slit J, are automatically accounted for and no calibration of the wedges is needed.

In the present model the number of terms of the Fourier series is restricted to 9 sine and 9 cosine components, as it was considered unnecessary to go any further in a first attempt. It is obviously also not possible to operate with negative coefficients as the light intensity is always positive. This serious limitation to the practical application of the instrument could be easily remedied by using a disk with four

rings for each harmonic, corresponding to $\sin nx$, $\cos nx$, $-\sin nx$, $-\cos nx$, instead of the two rings as in the present model. By preparing a larger disk or one with narrower rings this and an extension of the number of components up to the 20th harmonic could be achieved without much difficulty.

In a similar manner additional film disks corresponding to sets of different orthogonal functions could be easily prepared. By replacing the "harmonic" disk by another one of these, which could be done in a couple of minutes, the instrument would be ready to be used for the synthesis of a function with respect to the new set of component functions. The synthesis of Legendre functions in particular might be of some practical importance as it is equivalent to the building up of a function by a power series.

The figures 1-11 on fig. 5 give evidence of the proper functioning of the present instrument within the limitations mentioned. They are photographs of the trace on the oscillograph screen taken with a Zeiss Contax camera with a 1 : 2 Sonnar lens (and an additional lens extension piece) on 35 mm. Kodak high speed orthochromatic cinema film with $\frac{1}{6}$ sec. exposure time (photographs 1 and 2 are enlargements). As it appears from fig. 5 the instrument seems to be of special value as a demonstration apparatus apart from possible applications to problems of pure and applied science; *e. g.*, the synthesis of a complicated wave and the alteration of its shape produced by alterations in the intensities and the phases of the components, especially the addition and cutting off of such components can be observed directly. The successive building up of a triangular or a square wave and the effect of cutting off one after the other of the components starting from the top or from the bottom, as shown in these pictures, can be demonstrated in a minute or so and provides a very instructive demonstration experiment.

§ 3.

The coefficients in the expansion of a periodic function $f(t)$ with the fundamental period T into a Fourier series

$$f(t) = \frac{a_0}{2} + a_1 \cos \frac{2\pi}{T} t + a_2 \cos \frac{4\pi}{T} t + \dots \\ + b_1 \sin \frac{2\pi}{T} t + b_2 \sin \frac{4\pi}{T} t + \dots \quad (1)$$

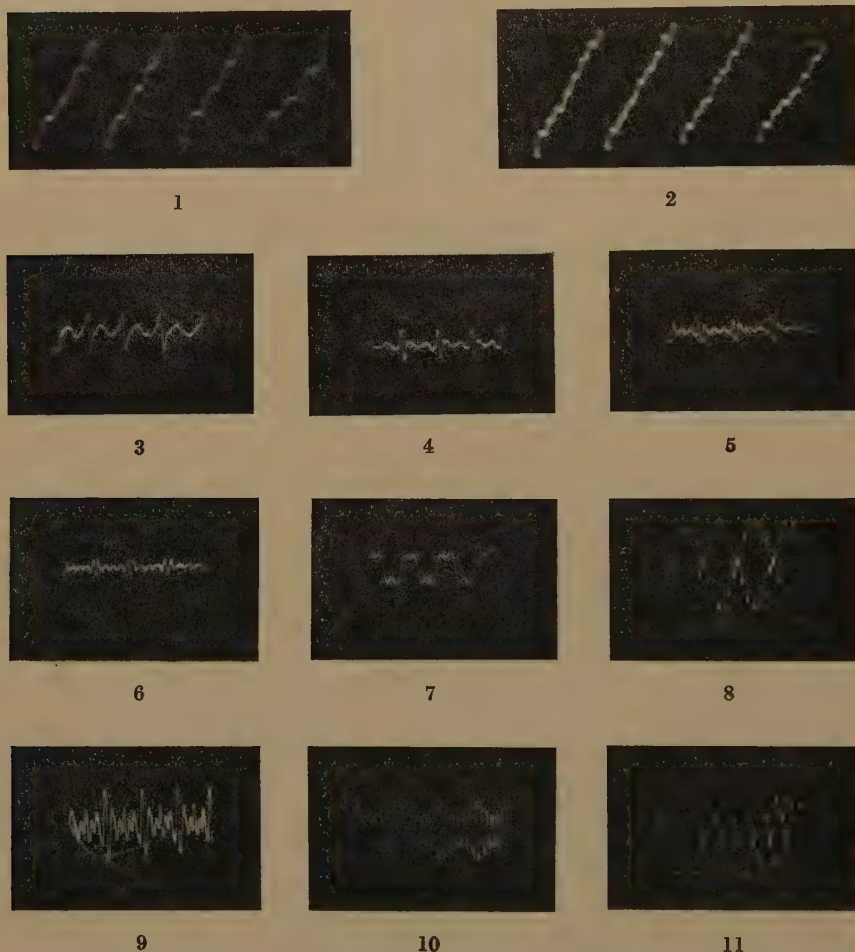
are given by the well-known formulæ

$$\left. \begin{aligned} a_n &= \frac{2}{T} \int_0^T \cos \frac{2\pi n}{T} t \cdot f(t) dt, \\ b_n &= \frac{2}{T} \int_0^T \sin \frac{2\pi n}{T} t \cdot f(t) dt. \end{aligned} \right\} \dots \quad (2)$$

Hence all methods for Fourier Analysis whose aim it is to calculate not only the amplitudes but also the phases of the harmonic components

must be based on some method for evaluating integrals of the form (2), *i. e.* integrals of the products of two periodic functions over the fundamental period. This can be achieved either numerically or graphically or by mechanical, electrical or optical means. The present method is based upon a principle which apparently has not been used before and which has been suggested to the authors by Professor M. Born.

Fig. 5.



Oscillographs obtained with the synthesiser. (See opposite page.)

It makes use of a well-known property of multi-grid valves commonly applied in the "mixer" valves of wireless superheterodyne receivers. If in such a valve, *e. g.*, an "octode" of the type shown in fig. 6, the grids 3, 4, 6, 7 are kept at constant potentials the number of electrons penetrating through the grid 4 from the cathode 1 depends entirely on the potential g_1 of the grid 2, and that linearly within the linear part of

the characteristic. Similarly, the fraction of these electrons collected by the anode 8 depends entirely on the potential g_2 of the grid 5 and that again linearly within a limited range. Hence the anode current i will be given by an expression of the form

$$i = (\alpha_1 + \beta_1 g_1)(\alpha_2 + \beta_2 g_2) = i_0 + Ag_1 + Bg_2 + Cg_1 g_2, \quad (3)$$

where i_0 , A, B, C are constants. The anode current therefore consists of four parts, one of which is constant and called i_0 , two of which are proportional to the two grid potentials resp., and a last one which is proportional to the product of the two grid potentials.

If now g_1 and g_2 are made to vary periodically in time with sufficiently high frequencies in such a way that their time averages vanish :

$$\bar{g}_1 = 0, \quad \bar{g}_2 = 0, \quad (4)$$

the value of the anode current, measured by a direct current galvanometer, will be equal to the time average \bar{i} of i which according to (3) is given by

$$\bar{i} - i_0 = C \frac{1}{T} \int_0^T g_1(t) g_2(t) dt, \quad (5)$$

where C is a constant of the circuit. The expression on the right-hand side of (5) is just an integral of the desired form. If therefore $g_1(t)$ is made to represent the function $f(t)$ to be analysed, and $g_2(t)$ one of the functions $\cos \frac{2\pi n}{T} t$ or $\sin \frac{2\pi n}{T} t$ the observation of \bar{i} will give directly the values of the Fourier coefficients according to (2).

Explanation of Fig. 5.

1. The first four terms in the Fourier series for a "saw-tooth" wave form, *i. e.*,

$$S_4(x) = \frac{2}{\pi} \sum_{n=1}^4 \frac{1}{n} \sin nx.$$

2. The first eight terms in the same Fourier series, *i. e.*,

$$S_8(x) = \frac{2}{\pi} \sum_{n=1}^8 \frac{1}{n} \sin nx.$$

3. $S_8(x) - S_4(x)$.

4. $S_8(x) - S_2(x)$.

5. $S_8(x) - S_3(x)$.

6. $S_8(x) - S_4(x)$.

7. The first two terms in the Fourier series for a "square wave," *i. e.*,

$$\frac{4}{\pi} \left(\sin x + \frac{1}{3} \sin 3x \right).$$

8. The production of "beats" by the synthesis of $\sin 6x + \cos 7x$.

9. The wave form of a "major chord" produced by the synthesis of $\sin 4x + \sin 5x + \sin 6x + \sin 8x$.

10. "Amplitude modulated" wave composed of "carrier" and two "side-bands":

$$0.3 \sin 5x + \sin 6x + 0.3 \sin 7x.$$

11. The wave form of a "major fifth" consisting of $\sin 4x + \sin 6x$.

This principle is, of course, by no means restricted to the Harmonic Analysis as the formulæ (2) and (5) hold quite generally for any set of orthogonal functions.

In order to apply the principle practically some means have to be found to produce the variable potentials g_1 and g_2 so as to represent accurately the given function $f(t)$ and the harmonic components. In the arrangement described in § 2 we already have a device for solving the latter problem. In order to produce potential variations proportional to $f(t)$ a suitable variable area drawing of this function, also in polar co-ordinates, is rotated together with the disk A bearing the harmonic functions, and a further light beam modulated by this pattern is concentrated on the cathode of another photocell. The electric outputs of the two photocells are then applied to the two control grids of a mixer valve of the described type and the anode current measured by a D.C. galvanometer. All the filters are set to zero transparency except one, which is set to maximum transparency, and the deflection of the galvanometer measured, from which the corresponding Fourier coefficient can be calculated. The same process is repeated for all the components in turn.

The graph of the function to be analysed is drawn in black on a circular thin celluloid disk of 2.6 cm. radius (or cut out of black paper) so that it fits exactly into the central part of the film disk A mentioned in § 2. It is set in the correct phase relationship with respect to the fundamental harmonic and in this position fixed and kept secured by the two glass plates and the locknut D. The insertion of the "function" or its replacement by another one takes only a couple of minutes.

The optical arrangement is essentially the same as that shown in fig. 1 except that instead of the single projecting lens S there are now used two similar lenses S_1 and S_2 , mounted one above the other, the top one for concentrating the light from the ring system on the cathode of one photocell T_1 as before, the bottom one for directing the beam through the central part of the disk on to the cathode of another photocell T_2 . The two cells are mounted inside a metal case U' (which replaces the case U used for the synthesis), which also contains the mixer valve, the batteries, and other parts of the circuit.

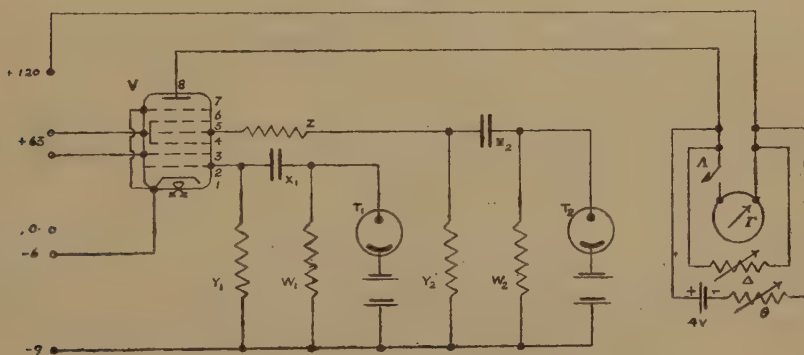
The electric circuit diagram is shown in fig. 6. The valve V is an indirectly heated octode valve (Mullard FC4). Preliminary experiments had been previously carried out with this valve in order to find out the most suitable conditions (*i. e.*, the values of the screen grid voltages and control grid biases) for using it as a multiplier device. These experiments had shown that under such conditions the valve had indeed the required properties and that the multiplication effect to be expected, *i. e.*, the increase $\delta \bar{i}$ in anode current caused by the application of oscillating grid potentials g_1, g_2 with time average $\delta \bar{s} = (\overline{g_1 g_2})$ was about

$$\delta \bar{i} / \delta \bar{s} = 200 \mu\text{A} / (\text{volt } R_{\mu s})^2.$$

The voltages for the two photocells T_1, T_2 are supplied by two separate

high tension batteries. As the cells have to be operated near the critical point (mentioned in § 2) for obtaining the highest sensitivity, these voltages must be maintained within a narrow margin of a few volts and are generally different for different individual cells. The grid potentials g_1, g_2 are developed across the load resistances W_1, W_2 of 15 megohms each, which are still small compared with the internal resistances of the cells but large enough for producing potential changes of about $\frac{1}{4}$ volt by the modulation of the light beams. These resistances are coupled to the control grids of V through the coupling condensers X_1, X_2 of 0.1 microfarad each in combination with the resistances Y_1, Y_2 of 10 megohms each in order to enforce the condition (4). As the impedance of these condensers is small compared with the resistances W and Y , even for the lowest frequencies used, no distortion of the waves with respect to amplitude or phase is introduced.

Fig. 6.



Circuit diagram of the analyser.

The relaxation time of the coupling systems is about 3 sec. Hence a time interval of about 10 sec. has to be inserted after any change in the illuminating system before readings of the anode current can be taken. It also ought to be mentioned that great care must be taken in providing proper insulation for all parts of the circuit carrying high tension in order to make the leakage resistances large compared with those of W and Y . A further safeguard is the insertion of the resistance Z (of 300 ohms) in one of the leads to the control grids of V in order to suppress the generation of parasitic high-frequency oscillations due to stray capacities and inductances in the circuit.

The anode current changes $\delta \bar{i}$ are measured by the balanced microammeter I (Unipivot galvanometer, sensitivity 5 div./ $\mu A.$, range $24 \mu A.$, internal resistance 350 ohms), the sensitivity of which can, if necessary, be reduced by the shunt A (10,000 ohm decade resistance box) and which can be disconnected by the key K . The balancing circuit consists of another 10,000 ohm decade resistance box B in connection with a 4 volt accumulator battery. Because of the extreme sensitivity of this

arrangement even a very small change in the heater current of the valve will produce a slow drift of the anode current so that the balancing has to be readjusted from time to time. Although the heater was always left on for about half an hour before taking readings a steady slow drift could not be avoided, as only slightly faulty accumulators were available. To this was added a certain amount of irregular fluctuation caused by the photocells operating near the critical point (as mentioned in § 2); but this fluctuation was only in the order of 1 division if the voltages were properly chosen so that it did not seriously affect the accuracy.

In principle, according to (5), it ought to be possible to determine a particular Fourier coefficient by opening the corresponding filter and then taking a reading of \bar{i} , and another one of i_0 after cutting off the light completely by covering the lenses S_1 and S_2 . In practice the procedure was a little more complicated because of the distortion due to the slight curvature of the valve characteristics. This effect is so small that it is completely negligible in the amplifying function of the valve. But as the coefficient of the product term on the right-hand side of (3) is small compared with the coefficients of the linear terms, even a very small "rectification" effect due to this distortion becomes comparable with the effect to be measured.

In order to eliminate this difficulty one has only to observe that equation (5) must now be replaced by the equation

$$\bar{i} = i_0 + (\overline{Ag_1}) + (\overline{Bg_2}) + C(\overline{g_1g_2}) \quad . \quad . \quad . \quad . \quad . \quad (6)$$

as A and B cannot now be regarded as constants; but equation (4) still holds. If now the beam affecting the potential g_1 is suppressed by covering the corresponding lens the current is changed to

$$\bar{i}' = i_0 + (\overline{Bg_2}), \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and, similarly, by suppressing the other beam the current becomes

$$\bar{i}'' = i_0 + (\overline{Ag_1}). \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Combining the equations (6), (7), (8) we have

$$i_0 - \bar{i}' - \bar{i}'' + \bar{i} = C(\overline{g_1g_2}) \quad . \quad . \quad . \quad . \quad . \quad (9)$$

instead of (5).

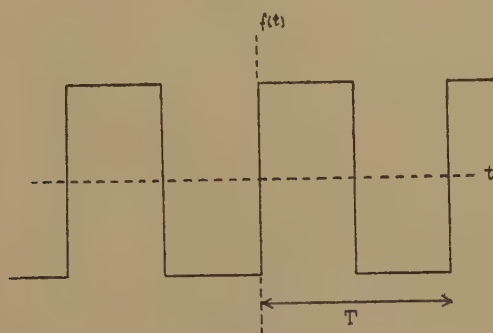
The distortion effect upon the term on the right-hand side of (9) can be neglected. Hence in order to determine the required Fourier coefficient one has to make four readings instead of two and to combine them as indicated by equation (9). In order to compensate the above mentioned steady drift in the anode current readings are always taken in the order indicated on the left-hand side of (9) in half-minute intervals. The instrument I must always be disconnected by the key A before making any changes in the illumination in order to safeguard it against the transient effects set up by such changes; it is reconnected 10 seconds before the next reading is taken.

In order to account for the lack of uniformity in the illumination of the different rings of the rotating disk correcting factors have to be

applied to the results obtained by means of formula (9) which are obviously inversely proportional to the average intensity of the corresponding light beams behind the rotating disk. These intensities can be very conveniently and accurately measured by using the instrument T (with a suitable shunt) as a ballistic galvanometer for measuring the total charge produced by the sudden uncovering (or sudden covering) of the lens S_1 . By this procedure also all deviations due to variations in the sensitivity of the photocell T_1 over the surface of its cathode are automatically accounted for.

The following preliminary measurements have been carried out for obtaining evidence for the practicability of the method :—

Fig. 7.



Square wave.

TABLE I.

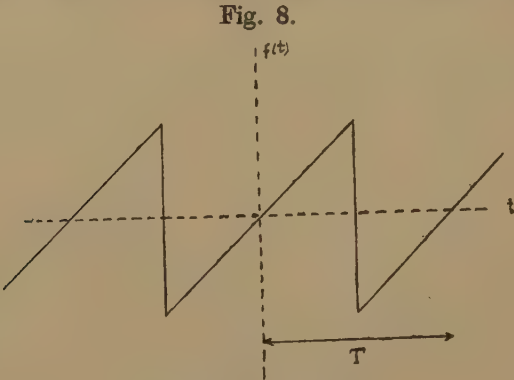
n .	a_n (exp.).	a_n (theor.).	b_n (exp.).	b_n (theor.).
1	0	0	42	42
2	0	0	1	0
3	0	0	13	14
4	0	0	0	0
5	1	0	7	8.4
6	-1	0	2	0
7	2	0	6	6

1. Analysis of the "square wave" represented by fig. 7. In the Fourier expansion of this function all the cosine terms vanish and also all the even sine terms; the coefficients of the odd sine terms are proportional to $1/n$. Table I. contains the results (in units of galvanometer deflections to the nearest whole number) up to the 7th harmonic which agree well with the theoretical values within the accuracy of observation (of the order of 1 division).

2. Analysis of the "triangular wave" represented by fig. 8. The Fourier expansion of this function also contains no cosine terms. The sine terms have alternative signs and the absolute values of the coefficients

are proportional to $1/n$. The results up to the 5th harmonic are given in Table II. and the agreement with the theoretical values is again fairly satisfactory.

These examples certainly give evidence for the validity of the method, although the accuracy achieved is rather low as compared with some other methods. But there is no doubt that a proper technical development of the instrument would greatly improve this and other deficiencies.



Triangular wave.

TABLE II.

n .	a_n (exp.).	a_n (theor.).	b_n (exp.).	b (theor.).
1	3	0	37	37
2	0	0	-19	-18.5
3	3	0	13	12.3
4	0	0	- 8	- 9.2
5	3	0	8	7.4

The authors wish to thank in the first place Professor M. Born for the suggestion of the principle of the analyser and for his interest in the work. They also wish to thank Professor C. G. Barkla and the members of his staff for putting some apparatus of the Physics Department at their disposal, and finally the University of Edinburgh for a grant from the Moray Fund for buying other pieces of apparatus.

List of References.

(1) L. W. Pollak, 'Rechentafeln zur harmonischen Analyse,' Leipzig, 1926.
L. W. Pollak, 'Handweiser zur harmonischen Analyse,' Prague, 1928.
C. A. Beevers and H. Lipson, Proc. Phys. Soc. xlviii. p. 772 (1936);
'Nature,' cxxxvii. p. 825 (1936).
M. A. S. Ross, 'Nature,' clii. p. 302, (1943).
(2) K. Bourne, 'Electronic Engineering,' xv. p. 149 (1942).
(3) L. J. Comrie, Journ. Sc. Instr. xxi. p. 129 (1944).

Department of Mathematical Physics,
University of Edinburgh.

LXXVII. *A Note on a Certain Multiple Integral.*

By A. S. GLADWIN, B.Sc., A.R.T.C., A.M.I.E.E.

(Communication from the Staff of the Research Laboratories of
The General Electric Co., Ltd., England *.)

[Received June 12, 1944.]

THE integral $I_n(x, y)$ is defined by the equation

$$I_n(x, y) = \int_x^y f(u_1) du_1 \int_{u_1}^y f(u_2) du_2 \dots \int_{u_{n-1}}^y f(u_n) du_n, \quad \dots \quad (1)$$

where $f(u)$ is a real single-valued function of a real variable u , and is continuous and finite in the closed interval $x \leq u \leq y$.

It will be shown that

$$I_n(x, y) = [I_1(x, y)]^n / n!,$$

where

$$I_1(x, y) = \int_x^y f(u) du,$$

From (1) it is clear that n can be only a positive integer. It will also be shown that

$$I_n(x, y) = \int_x^y f(u_1) du_1 \int_x^{u_1} f(u_2) du_2 \dots \int_x^{u_{n-1}^{n-1}} f(u_n) du_n.$$

From the definition of $f(u)$

$$\int_x^y f(u) du = \lim_{p \rightarrow \infty} \sum_{m=1}^p f(x + m \delta u) \delta u,$$

where

$$y - x = p \delta u.$$

$$\therefore [I_1(x, y)]^n = \left[\lim_{p \rightarrow \infty} \{f(x + \delta u) \delta u + \dots + f(y) \delta u\} \right]^n.$$

By the multinomial theorem

$$\begin{aligned} & [f(x + \delta u) \delta u + \dots + f(y) \delta u]^n \\ &= n! \sum \frac{\{f(x + \delta u) \delta u\}^{\alpha_1} \{f(x + 2\delta u) \delta u\}^{\alpha_2} \dots \{f(y) \delta u\}^{\alpha_p}}{\alpha_1! \alpha_2! \dots \alpha_p!}. \end{aligned} \quad (2)$$

The summation is carried out over all the positive integral values, including the value 0, which $\alpha_1, \alpha_2, \dots, \alpha_p$ may assume consistent with the relation

$$\sum_{m=1}^p \alpha_m = n,$$

* Communicated by Dr. C. C. Paterson, F.R.S.

and the theorem is true for all positive integral values of n , and for any finite number of terms, provided each term is finite and definite. The theorem is also true for an infinite number of terms provided the series

$$\sum_{m=1}^{\infty} f(x+m\delta u)\delta u$$

is absolutely convergent, as it is in this case (from the definition of $f(u)$). This result is easily obtained by repeated application of Cauchy's theorem concerning the multiplication of absolutely convergent series; for the left-hand side of (2) is simply the series

$$\sum_{m=1}^{\infty} f(x+m\delta u)\delta u$$

multiplied by itself $n-1$ times.

The limit of the left-hand side of (2) as $p \rightarrow \infty$ is $[I_1(x, y)]^n$. It follows from the above result that the limit of the right-hand side is also $[I_1(x, y)]^n$.

Consider, first, the case in which n of the α 's are $=1$, and the remaining $p-n$ are 0. The corresponding term is $n!$ times the sum of all the combinations of products of the factors $f(x+\delta u)\delta u \dots f(y)\delta u$ taken n at a time. Let $f(u_{n-1})\delta u$ denote a factor such that $x+(n-2)\delta u < u_{n-1} < y$.

The sum of the products of $f(u_{n-1})\delta u$ with each of the similar factors, denoted generally by $f(u_n)\delta u$, in the range $u_{n-1} < u_n \leq y$, is

$$f(u_{n-1})\delta u \sum_{u_n=u_{n-1}+\delta u}^y f(u_n)\delta u.$$

Similarly, if $f(u_{n-2})\delta u$ is another factor, such that

$$x+(n-3)\delta u < u_{n-2} < u_{n-1} < u_n \leq y,$$

the sum of the combinations of $f(u_{n-2})\delta u$ with each of the similar factors, taken two at a time, between $f(u_{n-2})\delta u$ and $f(y)\delta u$ (including $f(y)\delta u$), is

$$f(u_{n-2})\delta u \sum_{u_{n-1}=u_{n-2}+\delta u}^{y-\delta u} \left\{ f(u_{n-1})\delta u \sum_{u_n=u_{n-1}+\delta u}^y f(u_n)\delta u \right\}.$$

Continuing, the sum of the combinations of the products of the factor $f(u_1)\delta u$ with each of the similar factors, taken $(n-1)$ at a time, between $f(u_1)\delta u$ and $f(y)\delta u$ (including $f(y)\delta u$), is

$$f(u_1)\delta u \sum_{u_2=u_1+\delta u}^{y-(n-2)\delta u} \left\{ f(u_2)\delta u \sum_{u_3=u_2+\delta u}^{y-(n-3)\delta u} \left\{ f(u_3)\delta u \dots \sum_{u_n=u_{n-1}+\delta u}^y \{f(u_n)\delta u\} \dots \right\} \right\},$$

where

$$x < u_1 < \dots < u_n \leq y,$$

and the sum of all the combinations of all the factors, taken n at a time, in the range $x < u_1 < \dots < u_n \leq y$, is

$$\sum_{u_1=x+\delta u}^{y-(n-1)\delta u} \left\{ f(u_1)\delta u \sum_{u_2=u_1+\delta u}^{y-(n-2)\delta u} \left\{ f(u_2)\delta u \dots \sum_{u_n=u_{n-1}+\delta u}^y \{f(u_n)\delta u\} \dots \right\} \right\}.$$

As $p \rightarrow \infty$, $\delta u \rightarrow 0$, and the limit of each of the summations becomes a definite integral of finite value, which is a continuous function of its

limits. This follows from the definition of $f(u)$. The limit of the above expression is therefore

$$\int_x^y f(u_1) du_1 \int_{u_1}^y f(u_2) du_2 \dots \int_{u_{n-1}}^y f(u_n) du_n = I_n(x, y)$$

and the limit of the corresponding term in the expansion of (2) is $n! I_n(x, y)$. Consider next the case in which one of the α 's is 2, $(n-2)$ are 1, and the remainder are 0.

Proceeding as before, the corresponding term, S_1 , is seen to be

$$\begin{aligned} S_1 = & \frac{n!}{2} \sum_{u_2=x+\delta u}^{y-(n-2)\delta u} \left\{ f^2(u_2) \delta u^2 \sum_{u_3=u_2+\delta u}^{y-(n-3)\delta u} \left\{ f(u_3) \delta u \dots \sum_{u_n=u_{n-1}+\delta u}^y \{ f(u_n) \delta u \} \dots \right\} \right\} \\ & + \text{etc.} \\ & + \frac{n!}{2} \sum_{u_2=x+\delta u}^{y-(n-2)\delta u} \left\{ f(u_2) \delta u \sum_{u_3=u_2+\delta u}^{y-(n-3)\delta u} \left\{ f(u_3) \delta u \dots \right. \right. \\ & \quad \left. \left. \times \sum_{u_n=u_{n-1}+\delta u}^y \{ f^2(u_n) \delta u^2 \} \dots \right\} \right\}. \quad (3) \end{aligned}$$

If $f(u_2)$ is the greatest value of $f(u)$ in the range $x+\delta u \leq u \leq y$, then each of the above summations is less than

$$f(u_2) \delta u \frac{n!}{2} \sum_{u_2=x+\delta u}^{y-(n-2)\delta u} \left\{ f(u_2) \delta u \sum_{u_3=u_2+\delta u}^{y-(n-3)\delta u} \left\{ f(u_3) \delta u \dots \sum_{u_n=u_{n-1}+\delta u}^y \{ f(u_n) \delta u \} \dots \right\} \right\}.$$

This summation is finite for all finite values of n and for all values of p , so that each of the summations in (3) is less than $K\delta u$, K being finite, and the sum of the terms of (3) is less than

$$(n-1)K\delta u = K_{11}\delta u,$$

where K_{11} is finite for all finite values of n , and for all values of p :

$$S_1 < K_{11}\delta u.$$

Similarly, by choosing the minimum value of $f(u)$ it is obvious that $S_1 > K_{12}\delta u$, where K_{12} is finite in the same way as K_{11} :

$$K_{12}\delta u < S_1 < K_{11}\delta u.$$

Similarly, if two of the α 's are 2, or if one of the α 's is 3, and the remainder are 1 or 0, and the corresponding term in the expansion of (2) is S_2 , then

$$K_{22}\delta u^2 < S_2 < K_{21}\delta u^2.$$

K_{21} and K_{22} are also finite for all finite values of n and for all values of p .

Continuing the process over all the possible combinations of α 's, and denoting the sum of all the terms in the expansion of (2), other than $n! I_n(x, y)$, by S ,

$$K_{12}\delta u + K_{22}\delta u^2 + \dots + K_{r2}\delta u^r < S < K_{11}\delta u + K_{21}\delta u^2 + \dots + K_{r1}\delta u^r.$$

Let K_{q2} be the least of the quantities $K_{12} \dots K_{r2}$ and let K_{q1} be the greatest of the quantities $K_{11} \dots K_{r1}$, then

$$K_{q2}(\delta u + \delta u^2 + \dots + \delta u^r) < S < K_{q1}(\delta u + \delta u^2 + \dots + \delta u^r),$$

$$K_{q2}\delta u(1 - \delta u^r)/(1 - \delta u) < S < K_{q1}\delta u(1 - \delta u^r)/(1 - \delta u).$$

$$\therefore \quad \lim_{\delta u \rightarrow 0} S = 0.$$

It has been shown that the expression

$$\left[\sum_{m=1}^p f(x + m\delta u) \delta u \right]^n$$

may be expanded in a series, of which the limit of one term as $p \rightarrow \infty$ is $n! I_n(x, y)$, and of which the limit of the sum of all the remaining terms is zero :

$$[I_1(x, y)]^n = n! I_n(x, y),$$

$$\text{or} \quad I_n(x, y) = [I_1(x, y)]^n / n!$$

In finding the sum of the combinations of products of the various terms, the summations were taken between a given term and the upper limit y . But it is equally valid to consider the combinations obtained by starting at a given term and working always towards the lower limit x . Using exactly the same algebra, and altering only the limits of the summations (and integrals) it is easy to see that

$$\begin{aligned} [I_1(x, y)]^n / n! &= \int_x^y f(u_1) du_1 \int_x^{u_1} f(u_2) du_2 \dots \int_x^{u_{n-1}} f(u_n) du_n \\ &= I_n(x, y). \end{aligned}$$

LXXVIII. *Some Devices for the Solution of Large Sets of Simultaneous Linear Equations.*

(With an Appendix on the Reciprocation of Partitioned Matrices.)

By Prof. W. J. DUNCAN, D.Sc., M.I.Mech.E.*

[Received September 1, 1944.]

Summary.

The method proposed is not new in principle, but by help of the matrix notation it is expressed in the most general and concise form. The unknowns and equations are divided into subsets, which is equivalent to the partitioning of the matrices of the coefficients and constants. Operations with the sub-matrices yield the complete solution of the problem. The process is of particular value when facilities exist for the

* Communicated by the Author,

solution of a certain limited number of simultaneous linear equations and it is required to solve a more numerous set.

CONTENTS.

	Page
§ 1. Introduction	661
§ 2. Details of the Process	661
§ 3. The Calculation of Reciprocal Matrices	664
§ 4. Some Alternative Forms of the Reciprocal Matrix.	665
Appendix. Notes on the Reciprocation of Partitioned Matrices.....	667
List of References	670

§ 1. *Introduction.*

It sometimes happens that facilities exist for the solution of a given number n of simultaneous linear algebraic equations and that the need arises to solve a more numerous set. This problem can often be solved with sufficient accuracy by methods of successive approximation, but attention is drawn here to a method which, in theory at least, is completely accurate. This method is not new in principle, but it is hoped that the present exposition will be found helpful on account of its generality and conciseness. These qualities are secured by making use of the notation of matrices.

The method to be described here will be called the Method of Partial Solution or the Method of Sub-Matrices. It consists essentially in selecting a tractable number n of the equations and a set of n unknowns. The set of n selected equations is rewritten with the n selected unknowns on the left-hand side, while all the other terms are transferred to the right-hand side. The solution of this set is obtained in general form, so that the n selected unknowns are expressed explicitly as linear functions of the remaining unknowns. These values are substituted in the remaining equations, which then contain the remaining unknowns only. When these equations have been solved the first selected unknowns can be calculated at once.

Since this paper was circulated within the Aeronautical Research Committee the writer has learned that the method described in § 2 has been successfully used for some years by Mr. R. R. M. Mallock in conjunction with his electrical calculating machine, and that the expression for the reciprocal of a partitioned matrix which appears in equation (3.3) has been given by Dr. A. C. Aitken of Edinburgh in lectures to his students, together with some alternative equivalent forms which are now included in the present paper. Similar methods are described in a paper by Hotelling ⁽²⁾, which contains a useful bibliography.

§ 2. *Details of the Process.*

Suppose that the number of equations in the set is r and that a set of n equations is tractable with the facilities available. Let

$$r=n+m, \quad (2.1)$$

where m is of course positive. Then the complete set of equations can be written in the matrix form

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} g \\ h \end{bmatrix}, \quad \dots \quad (2.2)$$

where the orders * of the matrices are :—

$$A \quad (\bar{n} \ n)$$

$$B \quad (\bar{n} \ m)$$

$$C \quad (\bar{m} \ n)$$

$$D \quad (\bar{m} \ m)$$

$$x \quad (\bar{n} \ 1)$$

$$y \quad (\bar{m} \ 1)$$

$$g \quad (\bar{n} \ 1)$$

$$h \quad (\bar{m} \ 1).$$

When expanded, equation (2.2) yields

$$Ax + By = g \quad \dots \quad (2.3)$$

and

$$Cx + Dy = h. \quad \dots \quad (2.4)$$

It will be assumed that A is not singular†. Premultiply (2.3) by A^{-1} . Then

$$x = A^{-1}g - A^{-1}By. \quad \dots \quad (2.5)$$

When this is substituted in equation (2.4) the latter becomes

$$CA^{-1}g - CA^{-1}By + Dy = h,$$

or

$$Ey = h - CA^{-1}g, \quad \dots \quad (2.6)$$

with

$$E \equiv D - CA^{-1}B. \quad \dots \quad (2.7)$$

The matrix equation (2.6) represents a set of m scalar equations in the m unknowns y . When these have been solved, the unknowns x can be found at once from (2.5). If it should happen that m is intractably large the whole process must be applied to equation (2.6).

The process of computation ‡ can be set out as follows :—

Find in succession

$$(1) \ A^{-1}.$$

$$(2) \ A^{-1}B.$$

$$(3) \ CA^{-1}.$$

$$(4) \ CA^{-1}B.$$

$$(5) \ E \equiv D - CA^{-1}B = D - (4).$$

* The notation is that used in 'Elementary Matrices. (1)

† The selected equations and unknowns must be varied until a non-singular matrix A is obtained. This must be possible whenever the equations have a solution which is unique.

‡ For methods of calculating reciprocal matrices, see chap. iv. of ref. 1 and the Appendix to this paper.

$$(6) \quad CA^{-1}g.$$

$$(7) \quad y \text{ from } Ey = h - CA^{-1}g, \text{ where the right-hand side of the equation is completely known.}$$

$$(8) \quad x \text{ from } x = A^{-1}g - A^{-1}By.$$

A question of some importance is the best selection for the number n when r is given. When r is not greater than twice the number of equations which can be handled conveniently with the available aids to computation, it will probably be best to make n as small as possible or nearly so, since n is the order of the square matrix A which has to be reciprocated, and the calculation of a reciprocal is relatively laborious. The bounds to the number n will be this minimum and $r/2$ or $(r+1)/2$ according as r is even or odd.

Numerical Example.

This example is very simple and merely intended to illustrate the method. The equations to be solved are :—

$$z_1 + 2z_2 + z_3 + 3z_4 = 4,$$

$$2z_1 + 5z_2 + 2z_3 + z_4 = 10,$$

$$4z_1 + 3z_2 + 2z_3 + z_4 = 5,$$

$$z_1 + 2z_2 + 3z_3 + 4z_4 = 8.$$

$$\text{Put} \quad A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix},$$

$$g = \begin{bmatrix} 4 \\ 10 \end{bmatrix}, \quad \text{and} \quad h = \begin{bmatrix} 5 \\ 8 \end{bmatrix}.$$

$$(1) \quad A^{-1} = \begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix}.$$

$$(2) \quad A^{-1}B = \begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 13 \\ 0 & -5 \end{bmatrix}.$$

$$(3) \quad CA^{-1} = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 14 & -5 \\ 1 & 0 \end{bmatrix}.$$

$$(4) \quad CA^{-1}B = \begin{bmatrix} 14 & -5 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 37 \\ 1 & 3 \end{bmatrix}.$$

$$(5) \quad E = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} 4 & 37 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} -2 & -36 \\ 2 & 1 \end{bmatrix}.$$

$$(6) \quad CA^{-1}g = \begin{bmatrix} 14 & -5 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 10 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}.$$

$$(7) \quad Ey = h - CA^{-1}g,$$

$$\text{or} \quad -2z_3 - 36z_4 = -1$$

$$2z_3 + z_4 = 4.$$

These yield

$$z_3 = \frac{143}{70}, \quad z_4 = -\frac{6}{70}.$$

$$\begin{aligned}
 (8) \quad x &= A^{-1}g - A^{-1}By, \\
 &= \begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 10 \end{bmatrix} - \begin{bmatrix} 1 & 13 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} \frac{143}{70} \\ -\frac{6}{70} \end{bmatrix}, \\
 &= \frac{1}{70} \begin{bmatrix} -65 \\ 110 \end{bmatrix}.
 \end{aligned}$$

Finally, $z_1 = -\frac{65}{70}, \quad z_2 = \frac{110}{70}, \quad z_3 = \frac{143}{70}, \quad z_4 = -\frac{6}{70}.$

§ 3. The Calculation of Reciprocal Matrices.

Whenever the *general* solution of a set of linear algebraic equations having arbitrarily assignable numbers as "constants" but fixed coefficients is required, it becomes necessary to calculate the reciprocal of the matrix of the coefficients. This can be done by an extension of the method given in § 2, and the question is considered in a more general way in the Appendix.

By equation (2.6)

$$y = E^{-1}h - E^{-1}CA^{-1}g, \quad . \quad . \quad . \quad . \quad . \quad (3.1)$$

and therefore by (2.5)

$$x = (A^{-1} + A^{-1}BE^{-1}CA^{-1})g - A^{-1}BE^{-1}h. \quad . \quad . \quad . \quad (3.2)$$

From these equations it follows that the complete general solution of equations (2.3) and (2.4) is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A^{-1} + A^{-1}BE^{-1}CA^{-1} & -A^{-1}BE^{-1} \\ -E^{-1}CA^{-1} & E^{-1} \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix}, \quad . \quad . \quad (3.3)$$

where E is given by equation (2.7). Hence the square matrix on the right of (3.3) is the reciprocal of the matrix of the coefficients of (2.3) and (2.4), and it can be verified at once that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} A^{-1} + A^{-1}BE^{-1}CA^{-1} & -A^{-1}BE^{-1} \\ -E^{-1}CA^{-1} & E^{-1} \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_m \end{bmatrix} = I_r, \quad (3.4)$$

where I_s is the unit matrix of order s . An equivalent result is given on p. 113 of 'Elementary Matrices' ⁽¹⁾. The complete process of calculation is set out below, where the orders of the various matrices are given in brackets:—

Find

- (1) $A^{-1} \quad (\bar{n} \ n)$
- (2) $A^{-1}B \quad (\bar{n} \ m)$
- (3) $CA^{-1} \quad (\bar{m} \ n)$
- (4) $CA^{-1}B \quad (\bar{m} \ m)$
- (5) $E \equiv D - CA^{-1}B \quad (\bar{m} \ m)$
- (6) $E^{-1} \quad (\bar{m} \ m)$
- (7) $E^{-1}CA^{-1} = (6).(3) \cdot (\bar{m} \ n)$

$$\begin{aligned}(8) \quad & A^{-1}BE^{-1}=(2).(6) \quad (\bar{n} \ m) \\(9) \quad & A^{-1}BE^{-1}CA^{-1}=(8).(3)=(2).(7) \quad (\bar{n} \ n) \\(10) \quad & A^{-1}+A^{-1}BE^{-1}CA^{-1}=(1)+(9) \quad (\bar{n} \ n).\end{aligned}$$

The first five steps are identical with those given in § 2.

Numerical Example.

Find the reciprocal of the matrix of the coefficients in the set of equations occurring in the numerical example of § 2. Let this be partitioned into square sub-matrices of order 2, so that

$$A=\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}, \quad B=\begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}, \quad C=\begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}, \quad D=\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}.$$

The first five steps of the computation are as given in § 2 and the remainder are :—

$$(6) \quad E^{-1}=\frac{1}{70}\begin{bmatrix} 1 & 36 \\ -2 & -2 \end{bmatrix}.$$

$$(7) \quad E^{-1}CA^{-1}=\frac{1}{70}\begin{bmatrix} 1 & 36 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} 14 & -5 \\ 1 & 0 \end{bmatrix}=\frac{1}{70}\begin{bmatrix} 50 & -5 \\ -30 & 10 \end{bmatrix}.$$

$$(8) \quad A^{-1}BE^{-1}=\frac{1}{70}\begin{bmatrix} 1 & 13 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} 1 & 36 \\ -2 & -2 \end{bmatrix}=\frac{1}{70}\begin{bmatrix} -25 & 10 \\ 10 & 10 \end{bmatrix}.$$

$$(9) \quad A^{-1}BE^{-1}CA^{-1}=\frac{1}{70}\begin{bmatrix} -25 & 10 \\ 10 & 10 \end{bmatrix} \begin{bmatrix} 14 & -5 \\ 1 & 0 \end{bmatrix}=\frac{1}{70}\begin{bmatrix} -340 & 125 \\ 150 & -50 \end{bmatrix}.$$

$$\begin{aligned}(10) \quad & A^{-1}+A^{-1}BE^{-1}CA^{-1} \\ & =\begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix}+\frac{1}{70}\begin{bmatrix} -340 & 125 \\ 150 & -50 \end{bmatrix}=\frac{1}{70}\begin{bmatrix} 10 & -15 \\ 10 & 20 \end{bmatrix}.\end{aligned}$$

Finally,

$$\begin{bmatrix} 1 & 2 & 1 & 3 \\ 2 & 5 & 2 & 1 \\ 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}^{-1}=\frac{1}{70}\begin{bmatrix} 10 & -15 & 25 & -10 \\ 10 & 20 & -10 & -10 \\ -50 & 5 & 1 & 36 \\ 30 & -10 & -2 & -2 \end{bmatrix}.$$

§ 4. *Some Alternative Forms of the Reciprocal Matrix.*

Equation (2.4) yields

$$y=D^{-1}h-D^{-1}Cx,$$

and equation (2.3) therefore becomes

$$Fx=g-BD^{-1}h, \quad \dots \dots \dots (4.1)$$

$$\text{where} \quad F=A-BD^{-1}C. \quad \dots \dots \dots (4.2)$$

$$\text{Thus} \quad x=F^{-1}g-F^{-1}BD^{-1}h, \quad \dots \dots \dots (4.3)$$

$$\begin{aligned}\text{and} \quad & y=-D^{-1}CF^{-1}g \\ & +(D^{-1}+D^{-1}CF^{-1}BD^{-1})h. \quad \dots \dots \dots (4.4)\end{aligned}$$

Hence

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} F^{-1} & -F^{-1}BD^{-1} \\ -D^{-1}CF^{-1} & D^{-1} + D^{-1}CF^{-1}BD^{-1} \end{bmatrix}, \quad (4.5)$$

The identity of this with the reciprocal appearing in equation (3.3) can be proved directly as follows:—

From the definitions of E and F

$$EB^{-1}A = DB^{-1}A - C = DB^{-1}F, \quad (4.6)$$

and by reciprocation

$$A^{-1}BE^{-1} = F^{-1}BD^{-1}. \quad (4.7)$$

Also

$$\begin{aligned} DB^{-1}F &= (DB^{-1} - CA^{-1} + CA^{-1})F, \\ &= \{(D - CA^{-1}B)B^{-1} + CA^{-1}\}F, \\ &= (EB^{-1} + CA^{-1})F. \end{aligned}$$

Hence, by (4.6),

$$EB^{-1}A = (EB^{-1} + CA^{-1})F,$$

or

$$F^{-1} = A^{-1} + A^{-1}BE^{-1}CA^{-1}. \quad (4.8)$$

Similarly,

$$FC^{-1}D = AC^{-1}E,$$

from which it follows that

$$D^{-1}CF^{-1} = E^{-1}CA^{-1}. \quad (4.9)$$

and

$$E^{-1} = D^{-1} + D^{-1}CF^{-1}BD^{-1}. \quad (4.10)$$

It may be noted that any valid relation between the matrices can be transformed into another valid relation by the substitutions

$$A \rightarrow D$$

$$B \rightarrow C$$

$$C \rightarrow B$$

$$D \rightarrow A$$

which imply

$$E \rightarrow F$$

$$F \rightarrow E.$$

This follows from the fact that equation (2.2) can be written alternatively

$$\begin{bmatrix} D & C \\ B & A \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} h \\ g \end{bmatrix}. \quad (4.11)$$

On account of the relations established above, the reciprocal matrix can be written in the alternative forms

$$\begin{bmatrix} F^{-1} & -A^{-1}BE^{-1} \\ -D^{-1}CF^{-1} & E^{-1} \end{bmatrix}, \quad (4.12)$$

$$\begin{bmatrix} F^{-1} & -F^{-1}BD^{-1} \\ -E^{-1}CA^{-1} & E^{-1} \end{bmatrix}, \quad (4.13)$$

which were communicated to the writer by Dr. A. C. Aitken. It is possible that (4.5), (4.12) or (4.13) may have computational advantages in particular cases.

APPENDIX.

Notes on the Reciprocation of Partitioned Matrices.

As is well known, operations with partitioned matrices are (subject to the satisfaction of the usual conditions of conformability) conducted in the same way as for numerical matrices, *i. e.*, the sub-matrices are treated just as scalar elements, except that the order of the factors must be strictly maintained in all products. It follows from this that nearly all the processes for the computation of the reciprocals of numerical matrices can be adapted to partitioned matrices. Many methods for the calculation of reciprocals are described in chap. iv. of 'Elementary Matrices,' to which the reader may refer for further information.

One very easy process for finding the reciprocal of any square matrix u consists in performing identical "elementary operations" on the rows on the left of the two sides of the equation

$$u u^{-1} = I \quad \dots \dots \dots (1)$$

until u on the left side of the equation becomes reduced to I , at which stage I on the right is transformed into u^{-1} . These elementary operations on rows are equivalent to premultiplications by non-singular matrices, and therefore leave u^{-1} on the left-hand side of the equation unaltered throughout. The application of this method to a partitioned matrix will now be illustrated by a simple example.

Take the equation

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} u^{-1} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \quad \dots \dots \dots (2)$$

and premultiply both sides by $\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}$. The result is

$$\begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix} u^{-1} = \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \quad \dots \dots \dots (3)$$

For conciseness put

$$D - CA^{-1}B \equiv E, \quad \dots \dots \dots (4)$$

and premultiply by $\begin{bmatrix} A^{-1} & 0 \\ 0 & I \end{bmatrix}$. Then

$$\begin{bmatrix} I & A^{-1}B \\ 0 & E \end{bmatrix} u^{-1} = \begin{bmatrix} A^{-1} & 0 \\ -CA^{-1} & I \end{bmatrix} \quad \dots \dots \dots (5)$$

Successive premultiplications by

$$\begin{bmatrix} I & 0 \\ 0 & E^{-1} \end{bmatrix} \text{ and } \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \text{ yield}$$

$$\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} u^{-1} = u^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BE^{-1}CA^{-1} & -A^{-1}BE^{-1} \\ -E^{-1}CA^{-1} & E^{-1} \end{bmatrix} \quad \dots \dots (6)$$

This is identical with the reciprocal obtained in § 3. of the main text.

The explicit expression for x is found by postmultiplication by the reciprocal of the matrix coefficient of x . Then y can be obtained from (19) or (20) and, finally, z from one of the equations (16)–(18). Explicit formulæ for y and z can, of course, be found, but they will not be quoted.

By means of the above general solution it can be shown at once that the solution of the equations (10)–(12) is

$$a = G^{-1}(DG^{-1} - EH^{-1})^{-1} \begin{bmatrix} (AG^{-1} - BH^{-1})(DG^{-1} - EH^{-1})^{-1} \\ -(BH^{-1} - CJ^{-1})(EH^{-1} - FJ^{-1})^{-1} \end{bmatrix}^{-1}, \quad (22)$$

$$b = [G^{-1} - a(AG^{-1} - BH^{-1})] (DG^{-1} - EH^{-1})^{-1}, \quad (23)$$

$$c = (I - aA - bD)G^{-1}. \quad (24)$$

The matrices d , e and f are similarly found from the equations (10)–(12) with 0, I, 0 on the right instead of I, 0, 0. Lastly, g , h and j are found from the same equations with 0, 0, I on the right.

There is, of course, a connection between the solution of the matrix equations (13)–(15) and the solution of a set of three scalar simultaneous equations. When all the matrices appearing in (13)–(15) are of the first order they are permutable and behave as scalars. Hence (21) becomes

$$x \begin{bmatrix} \frac{l_1}{n_1} - \frac{l_2}{n_2} & \frac{l_2}{n_2} - \frac{l_3}{n_3} \\ \frac{m_1}{n_1} - \frac{m_2}{n_2} & \frac{m_2}{n_2} - \frac{m_3}{n_3} \end{bmatrix} = \begin{bmatrix} \frac{p_1}{n_1} - \frac{p_2}{n_2} & \frac{p_2}{n_2} - \frac{p_3}{n_3} \\ \frac{m_1}{n_1} - \frac{m_2}{n_2} & \frac{m_2}{n_2} - \frac{m_3}{n_3} \end{bmatrix}. \quad (25)$$

It can be shown that the coefficient of x is

$$\frac{n_2 \begin{vmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{vmatrix}}{(m_1 n_2 - m_2 n_1)(m_2 n_3 - m_3 n_2)},$$

while the expression on the right-hand side is

$$\frac{n_2 \begin{vmatrix} p_1 & m_1 & n_1 \\ p_2 & m_2 & n_2 \\ p_3 & m_3 & n_3 \end{vmatrix}}{(m_1 n_2 - m_2 n_1)(m_2 n_3 - m_3 n_2)}.$$

Thus the ordinary formula for x follows from equation (25).

On the subject of the solution of linear equations in non-commutative algebra the reader may consult A. R. Richardson, *Proc. Lond. Math. Soc.* vol xxviii. (ser. 2), p. 395 (1927).

Examples of the Reciprocation of a Partitioned Matrix.

(1) Consider the reciprocation of

$$\begin{bmatrix} f_{11}(u) & f_{12}(u) \\ f_{21}(u) & f_{22}(u) + k \end{bmatrix},$$

where the sub-matrices are, with the exception of D, all functions of a square matrix u . By the general formula (4) above

$$E = k + f_{22}(u) - f_{21}(u)f_{11}^{-1}(u)f_{12}(u),$$

Now functions of a given matrix are permutable, and this may therefore be written

$$E = k + f_{11}^{-1}(u)[f_{11}(u)f_{22}(u) - f_{12}(u)f_{21}(u)].$$

As a special case, suppose that the scalar determinant $\begin{vmatrix} f_{11}(x)f_{12}(x) \\ f_{21}(x)f_{22}(x) \end{vmatrix}$ is identically zero. Then also

$$f_{11}(u)f_{22}(u) - f_{12}(u)f_{21}(u) = 0,$$

so that

$$E = k.$$

It follows that the reciprocal matrix is

$$\begin{bmatrix} f_{11}^{-1}(u) + f_{11}^{-1}(u)f_{12}(u)k^{-1}f_{21}(u)f_{11}^{-1}(u) & -f_{11}^{-1}(u)f_{12}(u)k^{-1} \\ -k^{-1}f_{21}(u)f_{11}^{-1}(u) & k^{-1} \end{bmatrix}.$$

Evidently, then, the matrix is *singular* if k is nul or singular. It appears that similar theorems hold for matrices with any higher order of partitioning.

(2) Suppose that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

where the sub-matrices are all of the same order and therefore square. Then, if M is a non-singular matrix of the same order,

$$\begin{bmatrix} MA & MB \\ MC & MD \end{bmatrix}^{-1} = \begin{bmatrix} aM^{-1} & bM^{-1} \\ cM^{-1} & dM^{-1} \end{bmatrix}$$

and

$$\begin{bmatrix} AM & BM \\ CM & DM \end{bmatrix}^{-1} = \begin{bmatrix} M^{-1}a & M^{-1}b \\ M^{-1}c & M^{-1}d \end{bmatrix}.$$

These results can be extended for any order of partitioning.

References.

- (1) Frazer, Duncan and Collar, 'Elementary Matrices.' (Cambridge, 1938.)
- (2) Hotelling, "Some New Methods in Matrix Calculation," *Annals of Math. Statistics*, vol. xiv. p. 1 (March 1943).

LXXIX. *Expansions in Terms of Associated Legendre Functions.*

By T. M. MACROBERT (Glasgow) *.

[Received April 5, 1944.]

§ 1. *Introductory.*

THE main object of this paper is to show that a function of z , holomorphic in the ring-space bounded by two confocal ellipses with foci at $+1$ and

* Communicated by the Author.

-1 , can be expressed as the sum of two series, the first involving the functions $P_{m+n}^{-m}(z)$ dealt with in a recent paper*, and the second involving the corresponding functions $Q_{m+n}^m(z)$. This result will be found in formula (1) below. The particular case in which $m=0$ was given by K. Neumann†.

The proof is outlined in section 2, with a discussion of the validity of the proof in section 3. This involves an investigation, carried out in section 4, of the conditions under which $F(\alpha, \beta; \gamma; z)$ tends uniformly to unity when γ tends to infinity.

Particular cases, in which the second series does not appear, are dealt with in section 5, and a proof of these results for functions of a real variable satisfying Dirichlet's conditions is given in section 6.

An alternative discussion will be found in Szegő's 'Orthogonal Polynomials.' [American Mathematical Society, Colloquium Publications, vol. xxiii. (1939), Chapter IX.]

§ 2. Generalisation of an Expansion of Karl Neumann's.

It will now be proved that, if the function $f(z)$ is holomorphic within and on the boundary of the annulus bounded by confocal ellipses α and β in the z -plane with foci at 1 and -1 , such that α surrounds β , then at any interior point of the annulus

$$f(z) = \sum_{n=0}^{\infty} A_n (z^2 - 1)^{-\frac{1}{2}m} P_{m+n}^{-m}(z) + \sum_{n=0}^{\infty} B_n (z^2 - 1)^{\frac{1}{2}m} Q_{m+n}^m(z), \quad (1)$$

where $2\pi i A_n = (2m + 2n + 1) \int_{\alpha} f(\zeta) (\zeta^2 - 1)^{\frac{1}{2}m} Q_{m+n}^m(\zeta) d\zeta,$

and $2\pi i B_n = (2m + 2n + 1) \int_{\beta} f(\zeta) (\zeta^2 - 1)^{-\frac{1}{2}m} P_{m+n}^{-m}(\zeta) d\zeta.$

It should be noted that

$$\begin{aligned} (z^2 - 1)^{-\frac{1}{2}m} P_{m+n}^{-m}(z) &= (1 - z^2)^{-\frac{1}{2}m} T_{m+n}^{-m}(z) \\ &= \frac{1}{\Gamma(m+1) 2^m} F\left(\begin{matrix} -n, 2m+n+1 \\ m+1 \end{matrix}; \frac{1}{2} - \frac{1}{2}z\right), \end{aligned} \quad (2)$$

which is a polynomial in z , and that

$$\begin{aligned} Q_{m+n}^m(z) &= \frac{\Gamma(n+2m+1)\Gamma(\frac{1}{2})}{2^{n+m+1}\Gamma(n+m+\frac{3}{2})} \frac{(z^2-1)^{\frac{1}{2}m}}{z^{n+2m+1}} \\ &\quad \times F\left(\frac{1}{2}n+m+1, \frac{1}{2}n+m+\frac{1}{2}; n+m+\frac{3}{2}; z^{-2}\right), \end{aligned} \quad (3)$$

so that $(z^2 - 1)^{\frac{1}{2}m} Q_{m+n}^m(z)$ is uniform in the annulus if n is integral.

Now at any point z in the interior of the annulus

$$f(z) = \frac{1}{2\pi i} \int_{\alpha} \frac{f(\zeta)}{\zeta - z} d\zeta + \frac{1}{2\pi i} \int_{\beta} \frac{f(\zeta)}{z - \zeta} d\zeta.$$

* Quart. Journ. of Maths. Oxford Series, xiv. pp. 1-4 (1942).

† 'Ueber die Entwicklung einer funktion nach den Kugelfunktionen' (Halle, 1862).

In the second integral apply the formula *

$$\frac{1}{z-\zeta} = (z^2-1)^{\frac{1}{2}m} (\zeta^2-1)^{-\frac{1}{2}m} \sum_{n=0}^{\infty} (2m+2n+1) P_{m+n}^{-m}(\zeta) Q_{m+n}^m(z), \quad (4)$$

and in the first integral apply the same formula with z and ζ interchanged; then, on integrating term by term, the formula (1) is obtained.

It remains to prove that the term by term integration is justified.

Formula (4) may be established as follows. Multiply the recurrence formulæ

$$(2n+1)zQ_n^m(z) = (n-m+1)Q_{n+1}^m(z) + (n+m)Q_{n-1}^m(z), \quad (5)$$

$$(2n+1)\zeta P_n^{-m}(\zeta) = (n+m+1)P_{n+1}^{-m}(\zeta) + (n-m)P_{n-1}^{-m}(\zeta), \quad (6)$$

by $P_n^{-m}(\zeta)$ and $Q_n^m(z)$ respectively and subtract. Then, on replacing n by $m+n$ it is found that

$$\begin{aligned} & (2m+2n+1)(z-\zeta)P_{m+n}^{-m}(\zeta)Q_{m+n}^m(z) \\ &= \{(n+1)P_{m+n}^{-m}(\zeta)Q_{m+n+1}^m(z) - (2m+n+1)P_{m+n+1}^{-m}(\zeta)Q_{m+n}^m(z)\} \\ & \quad - \{nP_{m+n-1}^{-m}(\zeta)Q_{m+n}^m(z) - (2m+n)P_{m+n}^{-m}(\zeta)Q_{m+n-1}^m(z)\}. \end{aligned}$$

Here put $n=0, 1, 2, \dots, p$ in turn and add, noting that

$$P_m^{-m}(\zeta) = (\zeta^2-1)^{\frac{1}{2}m} 2^{-m} / \Gamma(m+1), \quad (7)$$

$$Q_m^m(z) = (z^2-1)^{-\frac{1}{2}m} 2^{m-1} \Gamma(m). \quad (8)$$

Thus

$$(\zeta^2-1)^{\frac{1}{2}m} (z^2-1)^{-\frac{1}{2}m} = (z-\zeta) \sum_{n=0}^p (2m+2n+1) P_{m+n}^{-m}(\zeta) Q_{m+n}^m(z) + R_p, \quad (9)$$

where

$$R_p = (2m+p+1)P_{m+p+1}^{-m}(\zeta)Q_{m+p}^m(z) - (p+1)P_{m+p}^{-m}(\zeta)Q_{m+p+1}^m(z). \quad (10)$$

Hence the formula (4) is valid if $R_p \rightarrow 0$ when $p \rightarrow \infty$, while the integrations are justified if the remainders $\rightarrow 0$ uniformly with respect to ζ in each case. These points will be dealt with in the two following sections.

§ 3. Discussion of the Remainder.

Here use will be made of the formulæ

$$\begin{aligned} Q_n^m(z) &= \sqrt{\left\{ \frac{\pi}{2\sqrt{z^2-1}} \right\}} \{z - \sqrt{z^2-1}\}^{n+\frac{1}{2}} \\ & \quad \times \frac{\Gamma(n+m+1)}{\Gamma(n+\frac{3}{2})} F\left\{ \begin{matrix} \frac{1}{2}+m, \frac{1}{2}-m; \\ n+\frac{3}{2} \end{matrix} ; \frac{-z+\sqrt{z^2-1}}{2\sqrt{z^2-1}} \right\}, \quad (11) \end{aligned}$$

$$\begin{aligned} P_n^m(z) &= e^{-\frac{1}{2}m\pi i} \frac{e^{\frac{1}{2}m\pi i}}{\sqrt{\{2\pi\sqrt{z^2-1}\}}} \frac{\Gamma(n+m+1)}{\Gamma(n+\frac{3}{2})} \\ & \quad \times \left[e^{(\frac{1}{2}-m)\frac{1}{2}\pi i} \{z - \sqrt{z^2-1}\}^{n+\frac{1}{2}} F\left\{ \begin{matrix} \frac{1}{2}+m, \frac{1}{2}-m; \\ n+\frac{3}{2} \end{matrix} ; \frac{-z+\sqrt{z^2-1}}{2\sqrt{z^2-1}} \right\} \right. \\ & \quad \left. + e^{-(\frac{1}{2}-m)\frac{1}{2}\pi i} \{z + \sqrt{z^2-1}\}^{n+\frac{1}{2}} F\left\{ \begin{matrix} \frac{1}{2}+m, \frac{1}{2}-m; \\ n+\frac{3}{2} \end{matrix} ; \frac{z+\sqrt{z^2-1}}{2\sqrt{z^2-1}} \right\} \right]. \quad (12) \end{aligned}$$

* Phil. Mag. ser. 7, xxi. p. 699 (1936).

Now it will be shown in the next section that, when $R(n) \rightarrow \infty$, $I(n)$ being bounded, the hypergeometric functions in these formulæ tend uniformly to unity for all points on the ellipses. Hence, in order to prove that the remainder in (4) tends uniformly to zero, we need only consider expressions such as

$$\{\zeta \pm \sqrt{(\zeta^2 - 1)}\}^{m+p+3/2} \{z - \sqrt{(z^2 - 1)}\}^{m+p+1/2} \quad \dots \quad (A)$$

and $\{\zeta \pm \sqrt{(\zeta^2 - 1)}\}^{m+p+1/2} \{z - \sqrt{(z^2 - 1)}\}^{m+p+3/2} \quad \dots \quad (B)$

the factors omitted being

$$\frac{\Gamma(p+2)\Gamma(2m+p+2)}{\Gamma(m+p+\frac{5}{2})\Gamma(m+p+\frac{3}{2})},$$

which tends to 1 when $p \rightarrow \infty$,

$$(z^2 - 1)^{-1/4+m/2} (\zeta^2 - 1)^{-1/4-m/2} (z - \zeta)^{-1},$$

the modulus of which is bounded when z and ζ lie on different fixed ellipses with ± 1 as foci, and constant factors.

Now let $z = \cosh w$, $\zeta = \cosh \omega$; then, if z and ζ lie on fixed ellipses with ± 1 as foci the real parts of w and ω are constant and may be taken to be positive. If ζ lies on the inner of the two ellipses, $R(w) > R(\omega)$, and the expressions (A) above become

$$e^{\pm(m+p+3/2)\omega - (m+p+1/2)w},$$

which tend uniformly to zero when $p \rightarrow \infty$.

The expressions (B) and the integral round the outer ellipse are dealt with in the same way.

§ 4. The Asymptotic Expansion of the Hypergeometric Function.

It has been shown * that, for large values of γ such that

$$|\operatorname{amp} \gamma| \leq \delta < \frac{1}{2}\pi, \quad F(\alpha, \beta; \gamma; z)$$

can be represented asymptotically by the first s terms of the hypergeometric series plus a remainder R_s , where

$$B(\beta+s, \gamma-\beta)R_s = T_{s+1} \int_0^1 \lambda^{\beta+s-1} (1-\lambda)^{\gamma-\beta-1} d\lambda \int_0^1 s(1-t)^{s-1} (1-\lambda tz)^{-\alpha-s} dt, \quad \dots \quad (13)$$

T_{s+1} being the $(s+1)$ th term of the series and s and $R(\gamma)$ being taken so large that $R(\alpha+s) > 0$, $R(\beta+s) > 0$, $R(\gamma-\beta) > 0$, provided that z is not real and ≥ 1 .

Consider the following region in the z -plane. Tangents are drawn from the origin to the circle $|z-1| = \sin \epsilon$, where ϵ is small and positive: these tangents make angles ϵ and $-\epsilon$ with the positive x -axis. Draw the circle $|z| = \kappa$, where κ is large and positive, and produce the two tangents to meet it. Then the region is bounded by the parts of the tangents between the circles, the smaller arc of the small circle and the

* Proc. Edin. Math. Soc. xli. p. 91 (1923).

larger arc of the large circle. In this region $|1-\lambda tz| \geq \sin \epsilon$, as it is the distance of the point 1 from the point λtz , where $0 \leq t \leq 1$, $0 \leq \lambda \leq 1$.

Thus, in this region $|(1-\lambda tz)^{-\alpha-s}| < M$, where M is a positive number independent of z and γ , and therefore, if $\beta = \sigma + i\tau$, $\gamma = \rho + i\mu$,

$$|R_s| \leq |T_{s+1}| \cdot M \cdot B(\sigma+s, \rho-\sigma) / B(\beta+s, \gamma-\beta).$$

Now, when ρ and γ tend to infinity,

$$\frac{\Gamma(\rho-\sigma)}{\Gamma(\rho+s)} \cdot \left| \frac{\Gamma(\gamma+s)}{\Gamma(\gamma-\beta)} \right| \rightarrow \mathcal{L} \frac{|\gamma^{\beta+s}|}{\rho^{\sigma+s}} = (\sec \chi)^{\sigma+s} e^{-\chi \tau},$$

where $\cos \chi = \mathcal{L} \rho / |\gamma|$.

Hence

$$|R_s| \leq |T_{s+1}| \cdot N,$$

where N is a positive number independent of $|\gamma|$ and of z in the region defined above.

Thus, when $|\gamma| \rightarrow \infty$, $|\text{amp } \gamma| \leq \delta < \frac{1}{2}\pi$, $F(\alpha, \beta; \gamma; z) \rightarrow 1$, the convergence being uniform in the region.

Note.—Since ϵ can be chosen as small as we please and κ as large as we please, it follows that $F(\alpha, \beta; \gamma; z) \rightarrow 1$ when $|\gamma| \rightarrow \infty$, $|\text{amp } \gamma| \leq \delta < \frac{1}{2}\pi$, for all finite values of z in the z -plane with a barrier along the real axis from 1 to $+\infty$. When $z=1$

$$F(\alpha, \beta; \gamma; 1) = \frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)},$$

$R(\gamma)$ being large, and this also tends to unity when $|\gamma| \rightarrow \infty$.

It will now be shown that, when ρ is large and μ is bounded, the asymptotic expansion also holds for z real and greater than unity; i. e., on the barrier approached from either side. To attain such a value, z must pass round the point 1 either below or above the point.

Regard λ and t as complex variables and in each of their planes let the paths of integration from 0 to 1 be deformed, when $|z| > 1$, into semi-circles on the diameters (0, 1) and lying below the real axes. Then if z , starting from a point in the straight line (0, 1) in the z -plane, passes by a path below the point 1 to a point for which $|z| > 1$ and $-\epsilon \leq \text{amp } z \leq 0$, $1-\lambda tz$ does not vanish* at any point of the path. Thus $|(1-\lambda tz)^{-\alpha-s}|$ is bounded and

$$\begin{aligned} |B(\beta+s, \gamma-\beta)R_s| &\leq |T_{s+1}| \cdot M \cdot \int_0^{\frac{\pi}{2}} (\cos \phi)^{\sigma+s-1} (\sin \phi)^{\rho-\sigma-1} d\phi \\ &\quad \times \int_0^{\frac{\pi}{2}} s (\sin \psi)^{s-1} d\psi, \end{aligned}$$

where $-\phi$ and $-\psi$ are the angles made by λ and t with the corresponding positive real axes, so that $\lambda = \cos \phi e^{-i\phi}$, $1-\lambda = \sin \phi e^{i(\frac{1}{2}\pi-\phi)}$, $t = \cos \psi e^{-i\psi}$, $1-t = \sin \psi e^{i(\frac{1}{2}\pi-\psi)}$, and M is a positive number independent of ρ .

* This point is dealt with more fully later on.

Hence

$$|B(\beta+s, \gamma-\beta)R_s| \leq |T_{s+1}| \cdot N \cdot B(\frac{1}{2}\sigma + \frac{1}{2}s, \frac{1}{2}\rho - \frac{1}{2}\sigma),$$

where N is a positive number independent of ρ .

But, when ρ tends to infinity, μ being bounded,

$$\left| \gamma^{-\frac{1}{2}\rho - \frac{1}{2}s} \frac{\Gamma(\gamma+s)}{\Gamma(\gamma-\beta)} \right| \times \frac{\Gamma(\frac{1}{2}\rho - \frac{1}{2}\sigma)}{\Gamma(\frac{1}{2}\rho + \frac{1}{2}s)} \rightarrow \mathcal{L} \frac{|\gamma^{\frac{1}{2}\rho + \frac{1}{2}s}|}{(\frac{1}{2}\rho)^{\frac{1}{2}\sigma + \frac{1}{2}s}} = 2^{\frac{1}{2}\sigma + \frac{1}{2}s}.$$

Therefore, when ρ is large,

$$|R_s| \leq |T_{s+1}| \gamma^{\frac{1}{2}\rho + \frac{1}{2}s} \times L,$$

where L is a positive number independent of ρ .

Thus, if $s > \sigma$, $R_s \rightarrow 0$ when $\rho \rightarrow \infty$, μ being bounded; consequently the asymptotic expansion holds in this region also.

If z approaches the upper side of the barrier the same result is obtained by taking the semi-circles above the real axes.

It follows that $F(\alpha, \beta; \gamma; z) \rightarrow 1$ when $\rho \rightarrow \infty$, μ being bounded, even when z is real and greater than 1.

It remains to investigate the conditions under which the convergence to the limit is uniform.

For this to be so $|1 - \lambda t z|$ must be greater than some positive number independent of z for all values of λ and t on the semi-circles and for z in some closed region.

Now in the case when the semi-circles are below the real axes

$$\lambda t z = \cos \phi e^{-i\phi} \cos \psi e^{-i\psi} |z| e^{-i\theta},$$

where $-\theta$ is the amplitude of z . Here

$$0 \leq \phi \leq \frac{1}{2}\pi, \quad 0 \leq \psi \leq \frac{1}{2}\pi,$$

and it will be assumed that $|z| > 1$, $0 \leq \theta \leq \epsilon$.

Then

$$\lambda t z = \cos \phi \cos \psi |z| \{ \cos(\phi + \psi + \theta) - i \sin(\phi + \psi + \theta) \},$$

so that

$$|1 - \lambda t z|^2 = \{1 - \cos \phi \cos \psi |z| \cos(\phi + \psi + \theta)\}^2 + \cos^2 \phi \cos^2 \psi |z|^2 \sin^2(\phi + \psi + \theta) \quad (A)$$

$$= \{\cos \phi \cos \psi |z| - \cos(\phi + \psi + \theta)\}^2 + \sin^2(\phi + \psi + \theta). \quad (B)$$

Hence, if $\phi \geq \frac{1}{4}\pi$, $\psi \geq \frac{1}{4}\pi$, $\cos(\phi + \psi + \theta)$ is negative or zero, and therefore, from (A),

$$|1 - \lambda t z| \geq 1.$$

Next let ϕ or ψ be greater than ϵ , ϕ and ψ not being both greater than $\frac{1}{4}\pi$; then

$$\sin(\phi + \psi + \theta) > \sin \epsilon,$$

and therefore, from (B)

$$|1 - \lambda t z| > \sin \epsilon.$$

Again, if ϕ and ψ are both $\leq \epsilon$, let $|z|$ be such that

$$|z| \cos^2 \epsilon \cos 3\epsilon \geq 1 + \sin \epsilon;$$

or

$$|z| \geq 1 / \{(1 - \sin \epsilon) \cos 3\epsilon\}.$$

Then $\cos \phi \cos \psi |z| \cos (\phi + \psi + \theta) \geq \cos^2 \epsilon \cdot |z| \cdot \cos 3\epsilon \geq 1 + \sin \epsilon$,
and therefore, from (A),

$$|1 - \lambda tz| \geq \sin \epsilon$$

Thus the convergence is uniform in the closed region bounded by the lines $\text{amp } z = 0$, $\text{amp } z = -\epsilon$, $|z| = 1/\{(1 - \sin \epsilon) \cos 3\epsilon\}$, $|z| = \kappa$, where κ is large.

Similarly, this holds for the corresponding region above the x -axis.

It follows that $F(\alpha, \beta; \gamma; z)$ tends uniformly to unity when $\rho \rightarrow \infty$, μ being bounded, in a closed region bounded by the circles $|z| = \kappa$, where κ is arbitrarily large, and $|z - 1| = r$, where r is chosen so that this circle encloses the three areas round $z = 1$ excluded from the regions of uniform convergence, and the part of the positive real axis between the circles. As ϵ is arbitrarily small r can be taken as small as we please.

Note.—If $\zeta = z/\sqrt{z^2 - 1}$, when z passes in the positive direction round an ellipse with ± 1 as foci from $\text{amp } z = 0$ to $\text{amp } z = \pi$, ζ passes completely round $\zeta = 1$ in the negative direction. The same is true when $\text{amp } z$ increases from π to 2π . By applying a simple transformation it can be seen that these results also hold when

$$\zeta = \frac{z + \sqrt{z^2 - 1}}{2\sqrt{z^2 - 1}}.$$

These paths in the ζ -plane can be included in a region in which $F(\alpha, \beta; \gamma; \zeta)$ tends uniformly to unity when $\rho \rightarrow \infty$, μ being bounded.

§ 5. *Some Particular Cases.*

If $f(z)$ is holomorphic at all points within and on the ellipse α , formula (1) reduces to

$$f(z) = \sum_{n=0}^{\infty} A_n (z^2 - 1)^{-\frac{1}{2}n} P_{m+n}^m(z), \quad \dots \dots \dots (14)$$

where
$$2\pi i A_n = (2m + 2n + 1) \int_{\alpha} f(\zeta) (\zeta^2 - 1)^{\frac{1}{2}m} Q_{m+n}^m(\zeta) d\zeta.$$

Now let the contour be deformed into the ξ -axis between -1 and 1 , described twice, and small circles about ± 1 . Then, if $R(m) > -1$, the integrals round the small circles tend to zero with their radii. This follows from the formulæ

$$Q_n^m(z) = \frac{1}{2} \Gamma(n + m + 1) \times \sum_{m, -m} \frac{\Gamma(m)}{\Gamma(n + m + 1)} \left(\frac{z + 1}{z - 1} \right)^{\frac{1}{2}m} F \left(\begin{matrix} -n, n + 1; \\ 1 - m \end{matrix} ; \frac{1}{2} - \frac{1}{2}z \right), \quad (15)$$

$$Q_n^m(z) = -\frac{1}{2} e^{-n\pi i} \Gamma(n + m + 1) \times \sum_{m, -m} \frac{\Gamma(m)}{\Gamma(n + m + 1)} \left(\frac{z - 1}{z + 1} \right)^{\frac{1}{2}m} F \left(\begin{matrix} -n, n + 1; \\ 1 - m \end{matrix} ; \frac{1}{2} + \frac{1}{2}z \right). \quad (16)$$

Hence, if $R(m) > -1$,

$$\begin{aligned} 2\pi i A_n &= (2m+2n+1) \int_{-1}^1 f(\xi)(1-\xi^2)^{\frac{1}{2}m} \left[\frac{-e^{\frac{1}{2}m\pi i} Q_{m+n}^m(\xi)}{+e^{-\frac{1}{2}m\pi i} Q_{m+n}^m(\xi, +1-)} \right] d\xi \\ &= i\pi \frac{\Gamma(n+2m+1)}{\Gamma(n+1)} (2m+2n+1) \int_{-1}^1 f(\xi)(1-\xi^2)^{\frac{1}{2}m} T_{m+n}^{-m}(\xi) d\xi, \end{aligned}$$

a result that follows from formula (15).

Thus, if z is holomorphic within and on α , and if $R(m) > -1$, [cf. formula (2)]

$$\begin{aligned} f(z) &= (1-z^2)^{-\frac{1}{2}m} \sum_{n=0}^{\infty} \frac{\Gamma(n+2m+1)}{n!} (m+n+\frac{1}{2}) T_{m+n}^{-m}(z) \\ &\quad \times \int_{-1}^1 f(\xi)(1-\xi^2)^{\frac{1}{2}m} T_{m+n}^{-m}(\xi) d\xi. \quad (17) \end{aligned}$$

If m is integral

$$\Gamma(n-m+1) T_n^m(z) = (-1)^m \Gamma(n+m+1) T_n^{-m}(z). \quad (18)$$

Therefore, if z is holomorphic within and on α , and if m is zero or a positive integer,

$$\begin{aligned} f(z) &= (-1)^m (1-z^2)^{-\frac{1}{2}m} \sum_{n=0}^{\infty} (m+n+\frac{1}{2}) T_{m+n}^{-m}(z) \\ &\quad \times \int_{-1}^1 f(\xi)(1-\xi^2)^{\frac{1}{2}m} T_{m+n}^{-m}(\xi) d\xi. \quad (19) \end{aligned}$$

In the statement of this theorem given in a former paper* the factor $(-1)^m$ was omitted.

§ 6. Proof based on Dirichlet Integrals.

In a former paper† formula (19) was established for functions satisfying Dirichlet's Conditions in the interval $(-1, 1)$. A corresponding proof will now be given for the more general formula (17).

The following formulæ, in which n , p and q are zero or positive integers, will be required:

$$\int_{-1}^1 T_{m+p}^{-m}(x) T_{m+q}^{-m}(x) dx = 0, \quad p \neq q, \quad R(m) > -1, \quad (20)$$

$$\int_{-1}^1 \{T_{m+n}^{-m}(x)\}^2 dx = \frac{2 \cdot n!}{\Gamma(2m+n+1) \cdot (2m+2n+1)}, \quad R(m) > -1. \quad (21)$$

In proving these use is made of the extended Rodrigues' formula

$$T_{m+n}^{-m}(z) = \frac{(-1)^n (1-z^2)^{-\frac{1}{2}m}}{2^{m+n} \Gamma(m+n+1)} \frac{d^n}{dz^n} \{(1-z^2)^{m+n}\}, \quad (22)$$

which can be established by writing

$$(1+z)^{m+n} = 2^{m+n} \{1 - (\frac{1}{2} - \frac{1}{2}z)\}^{m+n},$$

expanding by the binomial theorem, when $|\frac{1}{2} - \frac{1}{2}z| < 1$, and differentiating

* Phil. Mag. ser. 7, xxi. p. 699 (1936).

† Proc. Edin. Math. Soc. xlii. pp. 89-92 (1923).

term by term. In (20), if $p > q$, substitute from (22) for $T_{m+p}^{-m}(x)$ and from (2) for $T_{m+q}^{-m}(x)$ and integrate by parts $q+1$ times. Formula (21) can be derived in the same manner.

The summation formula

$$\sum_{n=0}^p \frac{\Gamma(n+2m+1)}{n!} (m+n+\frac{1}{2})(z-\zeta) T_{m+n}^{-m}(z) T_{m+n}^{-m}(\zeta) \\ = \frac{\Gamma(p+2m+2)}{2 \cdot p!} \{T_{m+p+1}^{-m}(z) T_{m+p}^{-m}(\zeta) - T_{m+p+1}^{-m}(\zeta) T_{m+p}^{-m}(z)\} \quad (23)$$

is also required. It can be deduced from the recurrence formula

$$(2n+1)z T_n^{-m}(z) = (n+m+1) T_{n+1}^{-m}(z) + (n-m) T_{n-1}^{-m}(z). \quad (24)$$

Let it be assumed that the expansion

$$f(x) = \sum_{n=0}^{\infty} A_n (1-x^2)^{-\frac{1}{2}m} T_{m+n}^{-m}(x), \quad (25)$$

where $R(m) > -1$, is valid for $-1 \leq x \leq 1$, and that it can be integrated term by term. Then, using the formulae (20) and (21), we see that

$$A_n = \frac{\Gamma(2m+n+1)}{n!} (m+n+\frac{1}{2}) \int_{-1}^1 (1-x^2)^{\frac{1}{2}m} f(x) T_{m+n}^{-m}(x) dx. \quad (26)$$

It will now be shown that, if $f(x)$ satisfies Dirichlet's Conditions, the series on the right of (25) converges to

$$\frac{1}{2} \{f(x+0) + f(x-0)\}$$

for $-1 < x < 1$.

Let Σ denote the sum of the first $n+1$ terms of the series; then, by (23),

$$(1-x^2)^{\frac{1}{2}m} \Sigma = \frac{\Gamma(n+2m+2)}{2 \cdot n!} \int_{-1}^1 (1-\xi^2)^{\frac{1}{2}m} f(\xi) \Omega d\xi,$$

where $\Omega = \{T_{m+n+1}^{-m}(x) T_{m+n}^{-m}(\xi) - T_{m+n+1}^{-m}(\xi) T_{m+n}^{-m}(x)\} / (x-\xi)$.

Now put $x = \cos \theta$, $\xi = \cos \phi$, substitute from the formula

$$T_n^{-m}(\cos \theta) = \frac{1}{\sqrt{(2\pi \sin \theta)}} \frac{\Gamma(n-m+1)}{\Gamma(n+\frac{3}{2})} \\ \times \sum_{i, -i} e^{(\frac{1}{2}+m)\frac{1}{2}\pi i - (n+\frac{1}{2})\theta i} F\left(\begin{matrix} \frac{1}{2}+m, \frac{1}{2}-m \\ n+\frac{3}{2} \end{matrix}; -\frac{e^{-i\theta}}{2i \sin \theta}\right), \quad (27)$$

and simplify. Then

$$(1-x^2)^{\frac{1}{2}m} \Sigma = \frac{1+k}{\pi} \int_0^\pi \frac{(\sin \phi)^{m+1}}{\sqrt{(\sin \theta \sin \phi)}} f(\cos \phi) \Delta d\phi,$$

where Δ denotes the expression

$$\left[\cos\{(\frac{1}{2}+m)\frac{1}{2}\pi - (n+m+\frac{3}{2})\theta\} \cos\{(\frac{1}{2}+m)\frac{1}{2}\pi - (n+m+\frac{1}{2})\phi\} \right. \\ \left. - \text{a similar expression with } \theta \text{ and } \phi \text{ interchanged} \right] \frac{1}{\cos \theta - \cos \phi} \\ = \frac{\sin\{(n+m+1)(\phi-\theta)\} \sin \frac{1}{2}(\phi+\theta) - \sin \frac{1}{2}(\phi-\theta) \cos\{(n+m+1)(\phi+\theta) - m\pi\}}{2 \sin \frac{1}{2}(\phi-\theta) \sin \frac{1}{2}(\phi+\theta)},$$

and $k \rightarrow 0$ when $n \rightarrow \infty$,

Hence, when $n \rightarrow \infty$,

$$(1-x^2)^{\frac{1}{2}m} \Sigma \rightarrow \frac{1}{2}(\sin \theta)^m [f\{\cos(\theta+0)\} + f\{\cos(\theta-0)\}],$$

or $\Sigma \rightarrow \frac{1}{2}\{f(x-0) + f(x+0)\},$

where $-1 < x < 1$.

As the formula (27) is not valid for $\theta=0$ or $\theta=\pi$, it is necessary, in order to complete the proof, to show that the integrals over the intervals $0 \leq \phi \leq \epsilon$ and $\pi - \epsilon \leq \phi \leq \pi$ tend to zero with ϵ .

If $R(m) > -\frac{1}{2}$, $R(n-m) > -1$,

$$\begin{aligned} & F\left(\frac{\frac{1}{2}-m, \frac{1}{2}+m}{n+\frac{3}{2}}; -\frac{e^{-i\phi}}{2i \sin \phi}\right) \\ &= \frac{1}{B(m+\frac{1}{2}, n-m+1)} \int_0^1 t^{m-\frac{1}{2}}(1-t)^{n-m} \left(1 + \frac{te^{-i\phi}}{2i \sin \phi}\right)^{m-\frac{1}{2}} dt. \end{aligned}$$

Now $|1 + te^{-i\phi}/(2i \sin \phi)| = |1 - \frac{1}{2}t + \frac{1}{2i}t \cot \phi| \geq \frac{1}{2},$

since $0 \leq t \leq 1$.

Hence, if $R(m) \leq \frac{1}{2}$,

$$| \{1 + te^{-i\phi}/(2i \sin \phi)\}^{m-\frac{1}{2}} | \leq 2^{\frac{1}{2}-R(m)} e^{\frac{1}{2}\pi \cdot |I(m)|}.$$

Thus, if $-\frac{1}{2} < R(m) \leq \frac{1}{2}$, $R(n-m) > -1$,

$$\left| F\left(\frac{\frac{1}{2}-m, \frac{1}{2}+m}{n+\frac{3}{2}}; -\frac{e^{-i\phi}}{2i \sin \phi}\right) \right| \leq M,$$

M being a positive number independent of ϕ and of $|n|$, where $|\text{amp } n| < \frac{1}{2}\pi$.

Again, if $0 \leq \phi \leq \pi$,

$$\sin \phi \cdot |1 + te^{-i\phi}/(2i \sin \phi)| = |\sin \phi(1 - \frac{1}{2}t) - \frac{1}{2}it \cos \phi| \leq \frac{3}{2},$$

the limit being taken as the value when $\phi=0$ or $\phi=\pi$.

Therefore, if $R(m) \geq \frac{1}{2}$,

$$|(\sin \phi)^{m-\frac{1}{2}} \{1 + te^{-i\phi}/(2i \sin \phi)\}^{m-\frac{1}{2}}| \leq \left(\frac{3}{2}\right)^{R(m)-\frac{1}{2}} e^{\frac{1}{2}\pi \cdot |I(m)|};$$

and, consequently, if $R(m) \geq \frac{1}{2}$, $R(n-m) > -1$,

$$\left| (\sin \phi)^{m-\frac{1}{2}} F\left(\frac{\frac{1}{2}-m, \frac{1}{2}+m}{n+\frac{3}{2}}; -\frac{e^{-i\phi}}{2i \sin \phi}\right) \right| \leq M,$$

M being a positive number independent of ϕ and of $|n|$, where $|\text{amp } n| < \frac{1}{2}\pi$.

Similar results hold for the conjugate function.

For the factors $|e^{\pm n\phi i}|$ in $|T_n^{-m}(\cos \phi)|$ to be bounded, $I(n)$ must be bounded.

Now, in the integral, n becomes $m+n$ or $m+n+1$, where n is a positive integer; thus $I(n)$ is bounded since m is fixed.

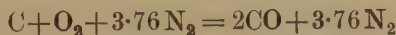
Hence, if $R(m) > -\frac{1}{2}$, and if n is large, the integrand tends to zero at both limits.

Again, as the hypergeometric function is symmetrical in m and $-m$, it follows, as above, that, if $R(-m) \geq \frac{1}{2}$, $R(n+m) > -1$,

$$\left| (\sin \phi)^{-m-\frac{1}{2}} F\left(\frac{\frac{1}{2}-m, \frac{1}{2}+m}{n+\frac{3}{2}}; -\frac{e^{-i\phi}}{2i \sin \phi}\right) \right| \leq M,$$

in which the two factors α and β express the degree of dissociation which has occurred. Equation (1), it should be noted, refers to combustion of the ideal air/fuel ratio mixture. Equations involving these two variables may, however, be written down for any air/fuel ratio.

It should be observed, also, that the form $C_xH_yO_zN_w$ may be applied also to a gaseous fuel mixture. Thus, suppose that the combustion phenomena of air-gas of the composition given by the right-hand side of the equation



were under consideration, then the form $C_xH_yO_zN_w$ would become $CON_{3.76}$, i. e., $x=1$, $y=1$, $z=0$ and $w=3.76$.

It is important to note, also, that when the nitrogen is not molecularly associated in the fuel, but is present as a constituent in a gas mixture, the total volume before combustion is always increased by w , if the gas is formed by decomposition with air into a mixture of CO, H_2 and N_2 . If other gases, say hydrocarbons, are present, then corresponding volume corrections are readily made by reference to the suitable modification of (1).

From (1) the ratio of the volume *after*, to the volume *before* combustion is given by

$$\sigma = \frac{x + 1.21y - 0.79z + 0.105(\alpha + \beta + w)}{x + y - z + 0.21}, \quad \dots \quad (2)$$

if the nitrogen is molecularly associated with the fuel. If the nitrogen is present as such, in a fuel gas mixture of CO, H_2 and N_2 , then

$$\sigma = \frac{x + 1.21y - 0.79z + 0.105(\alpha + \beta + w)}{x + y - z + 0.105(w + 2)}. \quad \dots \quad (2a)$$

2. First Relationship between α and the Temperature of Combustion.

This relationship is obtained by making use of the four well-known equations connecting the partial pressures and temperatures of the relevant gaseous mixtures.

The first set of these are the two following:—

$$\frac{P_{CO} \times P_{H_2O}}{P_{CO_2} \times P_{H_2}} = K, \quad \dots \quad (3)$$

with $\log_{10} K = A - \frac{B}{T_1} \quad \dots \quad (4)$

as a fairly close approximate expression connecting K with the absolute temperature T_1 of the mixture.

The other similar set of relationships is

$$\frac{(P_{CO})^2 \times P_{O_2}}{(P_{CO_2})^2} = K_1, \quad \dots \quad (5)$$

with $\log_{10} K_1 = A_1 - \frac{B_1}{T_1} \quad \dots \quad (6)$

Since the partial pressure of any constituent gas in a gaseous mixture is proportional to the number of molecules in or the relative volume of the constituent gas, then, from (1) and (3) it follows that

$$\beta = \frac{2y\alpha}{K(x-\alpha)+\alpha} \quad \dots \dots \dots (7)$$

Similarly from (5) $\beta = K_1 \left(\frac{x-\alpha}{\alpha} \right)^2 - \alpha \dots \dots \dots (7a)$

Combining (7) and (7a) gives

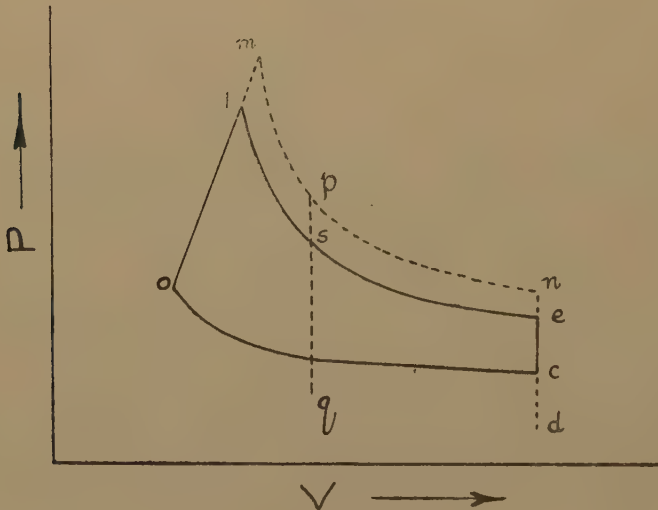
$$2y\alpha^3 = \{K(x-\alpha)+\alpha\} \{K_1(x-\alpha)^2 - \alpha^3\} \dots \dots \dots (8)$$

From this expression (8), a graph of α against T_1 is readily obtained.

3. *Second Relationship between α and the Temperature of Combustion.*

Suppose, now, that the combustion in an internal combustion engine follows the straight line "O1" in the PV diagram of fig. 1.

Fig. 1.



A straight combustion line is chosen for convenience, since it is analytically easier to evaluate, under variable specific heat conditions, either by calculation or by chart, both the corresponding energy and state changes in the gas, than it is to do so for lines following the polytropic type of change of state.

For such a straight combustion line it follows that, in general,

$$P = mV + n.$$

But

$$dq = K_v dT + PdV, \quad \dots \dots \dots (9)$$

$$\therefore dq = K_v dT + (mV + n)dV.$$

From $PV = RT$,
 is obtained $mV^2 + nV = RT$, (10)
 $\therefore 2mVdV + ndV = RdT$,

$$\therefore dV = \frac{RdT}{2mV+n},$$

$$\therefore dq = K_v dT + \frac{(mV+n)R}{2mV+n} dT. \quad \dots \quad (11)$$

From (10) $V = \frac{-n + \sqrt{n^2 + 4mRT}}{2m}.$

Therefore (11) becomes

$$\int_{Q_0}^Q dq = \int_{T_0}^T K_v dT + \int_{T_0}^T \frac{n + \sqrt{n^2 + 4mRT}}{2\sqrt{n^2 + 4mRT}} dT.$$

Taking $K_v = b + ST$,

this becomes $Q - Q_0 = (b + \frac{1}{2})(T - T_0) + \frac{S}{2}(T^2 - T_0^2)$
 $+ \frac{n}{4mR} \{ (n^2 + 4mRT)^{\frac{1}{2}} - (n^2 + 4mRT_0)^{\frac{1}{2}} \}. \quad \dots \quad (12)$

Referring now to fig. 1, another relation between α and T_1 may be obtained therefrom, as follows:—

The temperature T_1 at the peak of the full line diagram represents the true temperature under dissociation conditions, whereas the temperature T_m , indicated by the peak of the dotted extension line, represents the maximum temperature which would have been attained, had there been no dissociation. The point m can readily be determined by means of a variable specific heat internal combustion engine chart. The ratio

$$\phi_1 = \frac{\text{Heat developed under "dissociation" conditions from 0 to 1}}{\text{Heat developed under "no dissociation" conditions from 0 to } m}$$

may be expressed in two forms, as follows:—

The first is given by

$$\phi_1 = \int_{T_0}^{T_1} F(T) dT / \int_{T_0}^{T_m} F(T) dT, \quad \dots \quad (13)$$

where $F(T)$ is given by the right-hand side of (12).

The value of T_0 is obtained from chart, and so (13) can be evaluated by means of (12).

The other form of expression for ϕ_1 is

$$\phi_1 = \frac{6C_1(x - \alpha) + 6C_2\alpha + (2y - \beta)C_3}{6C_1x + 2C_3y}, \quad \dots \quad (14)$$

where C_1 is the heat of combustion of 1 lb. of carbon to carbon dioxide,
 C_2 is the heat of combustion of 1 lb. of carbon to carbon monoxide,
 C_3 is the heat of combustion of 1 lb. of hydrogen to water.

The foregoing expression (14) is obtained as follows :—

The gravimetric proportion of carbon in the fuel is given by

$$c = \frac{6x}{6x + 2y + 16z + 7w}, \quad (15)$$

while that of hydrogen is

$$h = \frac{2y}{6x + 2y + 16z + 7w}.$$

Since the denominator of (15) and of (16) is the same, it will not appear in the ratio expression (14), the derivation of which should then be clear.

Equating (13) and (14) and writing β in terms of α as given by (7a) gives

$$\int_{T_0}^{T_1} F(T) dT / \int_{T_0}^{T_m} F(T) dT = \frac{6C_1(x - \alpha) + 6C_2\alpha + \left\{ 2y - K_1 \left(\frac{x - \alpha}{\alpha} \right)^2 + \alpha \right\} C_3}{6C_1x + 2C_3y} \quad . . . (17)$$

as a second relation between α and T_1 .

By plotting the two graphs between α and T_1 , as obtained from (8) and (17), their intersection will give the true explosion temperature T_1 , the corresponding value of α and, therefore, of β .

4. *Determination of the True Expansion Line.*

To determine any point S on the true expansion line, *i. e.*, the line which includes the effect of re-association, the following procedure has been developed.

The dotted line “*mpn*,” which is obtained by chart, represents expansion under “no dissociation” conditions throughout. The point *q* on the constant volume line *pq* is determined by chart such that the heat developed from *q* to *p* is equal to that developed from *o* to *m*. At point S, the new values of α and β are now designated by α_1 and β_1 , for which the relationship given by (8) can be used, *i. e.*,

$$2y\alpha_1^3 = \{K(x - \alpha_1) + \alpha_1\} \{K_1(x - \alpha_1)^2 - \alpha_1^3\}. \quad . . . (18)$$

Also, as before, in the determination of (13), (14) and (17) the ratio

$$\phi_s = \frac{\text{Heat developed under “dissociation” conditions from } q \text{ to } S}{\text{Heat developed under “no dissociation” conditions from } q \text{ to } p}$$

may be expressed in two forms, as follows :—

$$\phi_s = \int_{T_q}^{T_s} F_1(T) dT / \int_{T_q}^{T_p} F_1(T) dT, \quad (19)$$

where, since *pq* is a constant volume line,

$$F_1(T) = K_v = b + ST.$$

Also, as before, in (14)

$$\phi_s = \frac{6C_1(x - \alpha_1) + 6C_2\alpha_1 + (2y - \beta_1)C_3}{6C_1x + 2C_3y}. \quad (20)$$

Hence, equating (19) and (20) gives

$$\frac{b(T_s - T_q) + \frac{S}{2}(T_s^2 - T_q^2)}{b(T_p - T_q) + \frac{S}{2}(T_p^2 - T_q^2)} = \frac{6C_1(x - \alpha_1) + 6C_2\alpha_1 + \left\{2y - K_1\left(\frac{x - \alpha_1}{\alpha_1}\right)^2 + \alpha_1\right\}C_3}{6C_1x + 2C_3y} \quad (21)$$

Equations (19) and (21) provide two graphs for different relationships between T_s and α_1 , the point of intersection of which gives the true value of T_s and also of α_1 and β_1 .

5. Conclusion.

By means of the foregoing analysis, it is possible, therefore, without undue computational labour, to obtain the true ideal engine indicator diagrams of any cycle and for any fuel, and so enable comparisons to be made between the behaviours of different fuels under dissociation and re-association conditions.

Such comparisons should provide useful information regarding the effects of dissociation and re-association on the power development obtained from different fuels at different compression ratios. It should also throw some valuable light on the hitherto unknown quantitative value of the re-associable heat discharged to exhaust, under varying compression ratio conditions for each fuel. Expressed as a fraction of the total heat supplied, this re-associable heat will be equal to $1 - \phi_e$, where ϕ_e is given by

$$\phi_e = \frac{\text{Heat developed under "dissociation" conditions from } d \text{ to } e}{\text{Heat developed under "no dissociation" conditions from } d \text{ to } n},$$

the heat developed from d to n being equal to that developed from o to m .

A CORRECTION.

In a recently published paper under a similar title by the same author (Phil. Mag. (ser. 7), xxxiv. pp. 486-488, 1943), the correction given below should be made:—

After equation (12) the following paragraph should have appeared in the manuscript:—

"A similar equation (13) can be derived in the same way by the relations

$$\frac{(P_{co})^2 \times P_o}{(P_{co})^2} = K_1,$$

and $\log_{10} K_1 = 8.46 - 26.600/T."$

In the paragraph which immediately follows equation (12), *for* "expression (12) can be readily solved graphically" *read* "expressions (12) and (13) can be readily solved graphically for T and α ."

LXXXI. *Sums of Series of the Form $\sum_{\alpha} a_n J_{n+\alpha}(z) J_{n+\beta}(z)$.*

By S. O. RICE*.

[Received August 9, 1943.]

1. *Introduction.*

IN a recent paper † S. A. Schelkunoff has expressed certain quantities associated with the impedance of an antenna as an infinite series involving Bessel functions. He also shows that a different physical argument leads to expressions involving Ci and Si functions for these same quantities.

It is of some interest, from a mathematical point of view, to be able to go directly from the series to the closed expression without appealing to the physics of the problem. A method of doing this is sketched here. It is rather involved and its existence probably would not have been suspected except for Schelkunoff's work. Additional results which were obtained in the search for the method are also stated.

The equations, obtained from Schelkunoff's paper, to be verified are

$$2\pi \sum_{m=0}^{\infty} \frac{4m+3}{(m+1)(2m+1)} J_{2m+3/2}^2(z) = 2(C + \log 2z - \text{Ci } 2z) \\ + (C + \log z - 2\text{Ci } 2z + \text{Ci } 4z) \cos 2z \\ + (\text{Si } 4z - 2\text{Si } 2z) \sin 2z, \quad \dots \quad (1.1)$$

$$-2\pi \sum_{m=0}^{\infty} \frac{4m+3}{(m+1)(2m+1)} J_{2m+3/2}(z) J_{-2m-3/2}(z) \\ = 2\text{Si } 2z + (\text{Ci } 4z - \log z - C) \sin 2z \\ - \text{Si } 4z \cos 2z, \quad \dots \quad (1.2)$$

where $C=577$ is Euler's constant and we have used

$$N_{2m+3/2}(z) = J_{-2m-3/2}(z).$$

2. *Verification of (1.1) and (1.2).*

When we use the result ‡

$$J_{\mu}(z) J_{\nu}(z) = \sum_{s=0}^{\infty} \frac{(-)^s (z/2)^{\mu+\nu+2s} \Gamma(\mu+\nu+2s+1)}{s! \Gamma(\mu+s+1) \Gamma(\nu+s+1) \Gamma(\mu+\nu+s+1)}, \quad (2.1)$$

* Communicated by the Author.

† "Antennas of Arbitrary Size and Shape," Proc. I.R.E. vol. xxix. pp. 493-521 (1941).

‡ G. N. Watson, 'Theory of Bessel Functions' (Camb. Univ. Press), p. 147.

we obtain

$$\left. \begin{aligned} \sum_{n=0}^{\infty} a_n J_{n+\alpha}(z) J_{n+\beta}(z) \\ = \sum_{t=0}^{\infty} \frac{(-)^t (z/2)^{2t+\alpha+\beta} \Gamma(2t+\alpha+\beta+1)}{t! \Gamma(t+\alpha+1) \Gamma(t+\beta+1) \Gamma(t+\alpha+\beta+1)} \\ \sum_{n=0}^{\infty} \frac{(-t)_n a_n}{(t+\alpha+\beta+1)_n} \end{aligned} \right\}, \quad (2.2)$$

$$\left. \begin{aligned} \sum_{n=0}^{\infty} a_n J_{n+\alpha}(z) J_{-n+\beta}(z) \\ = \sum_{t=0}^{\infty} \frac{(-)^t (z/2)^{2t+\alpha+\beta} \Gamma(2t+\alpha+\beta+1)}{t! \Gamma(t+\alpha+1) \Gamma(t+\beta+1) \Gamma(t+\alpha+\beta+1)} \\ \sum_{n=0}^{\infty} \frac{(-)^n (-\beta-t)_n a_n}{(\alpha+t+1)_n} \end{aligned} \right\}, \quad (2.3)$$

provided a_n is such that the series converge. We use the notation $(\alpha)_n = \alpha(\alpha+1) \dots (\alpha+n-1)$, $(\alpha)_0 = 1$. A number of results may be obtained directly from (2.2) and (2.3) by choosing a_n so that the series in n on the right have simple sums.

In order to establish (1.1) and (1.2) we set

$$\begin{aligned} a_0, a_2, a_4, \dots &= 0, \\ a_n &= a_{2m+1} = \frac{4m+3}{(m+1)(2m+1)}, \end{aligned}$$

$n=2m+1=\text{odd number}$,

$$\alpha=\beta=\frac{1}{2} \text{ in (2.2), } \alpha=\frac{1}{2}, \beta=-\frac{1}{2} \text{ in (2.3).}$$

We shall show later that

$$\begin{aligned} \sum_{m=0}^{\infty} \frac{(-\mu)_{2m+1}}{(\mu+2)_{2m+1}} \frac{4m+3}{(m+1)(2m+1)} &= - \int_0^1 \frac{1-x^\mu}{1-x} dx \\ &= -C - \psi(\mu+1), \quad (2.4) \end{aligned}$$

where $R(\mu) > -1$, so that the series and integral converge, C is Euler's constant and $\psi(\mu+1)$ is the logarithmic derivative of $\Gamma(\mu+1)$. Setting $\mu=t$ and $\mu=\beta+t=t-\frac{1}{2}$ for use in (2.2) and (2.3) respectively, and using (2.1) to sum the series in t gives

$$\begin{aligned} \sum_0^{\infty} \frac{4m+3}{(m+1)(2m+1)} J_{2m+3/2}^2(z) \\ = - \int_0^1 \frac{dx}{1-x} [J_\alpha(z) J_\beta(z) - x^{-\frac{\alpha+\beta}{2}} J_\alpha(2zx^{1/2}) J_\beta(2zx^{1/2})]_{\substack{\alpha=1/2 \\ \beta=1/2}} \\ = - \frac{2}{\pi z} \int_0^1 \frac{dx}{1-x} [\sin^2 z - x^{-1} \sin^2 zx^{1/2}], \quad (2.5) \\ \sum_0^{\infty} \frac{4m+3}{(m+1)(2m+1)} J_{2m+3/2}(z) J_{-2m-3/2}(z) \end{aligned}$$

$$\begin{aligned}
&= - \int_0^1 \frac{dx}{1-x} \left[-J_\alpha(z) J_\beta(z) + x^{-\frac{\alpha+\beta}{2} - \frac{1}{2}} J_\alpha(2zx^{1/2}) J_\beta(2zx^{1/2}) \right]_{\substack{\alpha=1/2 \\ \beta=-1/2}} \\
&= \frac{2}{\pi z} \int_0^1 \frac{dx}{1-x} [\sin z \cos z - x^{-1} \sin zx^{1/2} \cos zx^{1/2}], \quad \dots \quad (2.6)
\end{aligned}$$

where we have used

$$J_{1/2}(z) = \left(\frac{2}{\pi z}\right)^{1/2} \sin z, \quad J_{-1/2}(z) = \left(\frac{2}{\pi z}\right)^{1/2} \cos z. \quad \dots \quad (2.7)$$

Using $2 \sin^2 z = 1 - \cos 2z$, $2 \sin z \cos z = \sin 2z$, the last integrals in (2.5) and (2.6) become

$$\frac{1}{\pi z} \int_0^1 \frac{dx}{1-x} \left[\frac{1-x}{x} + \cos 2z - x^{-1} \cos 2zx^{1/2} \right]$$

and

$$\frac{1}{\pi z} \int_0^1 \frac{dx}{1-x} [\sin 2z - x^{-1} \sin 2zx^{1/2}],$$

respectively. These are the real and imaginary parts of

$$\frac{1}{\pi z} \int_0^1 \frac{dx}{x} \left[1 + \frac{1}{1-x} (xe^{i2z} - e^{i2zx^{1/2}}) \right].$$

Putting $x=y^2$ transforms this integral into

$$\frac{2}{\pi z} \int_0^1 \frac{dy}{y} \left[1 + \frac{1}{1-y^2} (y^2 e^{i2z} - e^{i2zy}) \right].$$

Using the partial fraction expansions

$$\frac{2y}{1-y^2} = \frac{1}{1-y} - \frac{1}{1+y},$$

$$\frac{2}{y(1-y^2)} = \frac{2}{y} + \frac{1}{1-y} - \frac{1}{1+y}$$

gives

$$\begin{aligned}
\frac{1}{\pi z} \int_0^1 \frac{dy}{y} \left[\frac{2-2e^{i2zy}}{y} + \frac{1}{1-y} (e^{i2z} - e^{i2zy}) \right. \\
\left. - \frac{1}{1+y} (e^{i2z} - e^{i2zy}) \right]. \quad \dots \quad (2.8)
\end{aligned}$$

The first part of the integrand may be evaluated by using

$$\left. \begin{aligned} C + \log x - Ci(x) &= \int_0^x \frac{dy}{y} (1 - \cos y), \\ Si(x) &= \int_0^x \frac{dy}{y} \sin y \end{aligned} \right\} \quad \dots \quad (2.9)$$

The remaining parts may be put into a similar form by changing the variable of integration. They yield

$$\begin{aligned}
&\frac{1}{\pi z} \int_0^1 \frac{du}{u} e^{i2z} (1 - e^{-2izu}) \\
&- \frac{1}{\pi z} \int_1^2 \frac{du}{u} (e^{i2z} - e^{-2iz + i2uz}).
\end{aligned}$$

The first term is of the same form as the first term of (2.8). The second term is

$$-\frac{1}{\pi z} (e^{i2z} - e^{-i2z}) \log 2 - \frac{e^{-i2z}}{\pi z} \int_1^2 \frac{du}{u} (1 - e^{i2uz}).$$

Collecting terms shows that (2.8) is equal to

$$\begin{aligned} & \frac{1}{\pi z} \left[(2 + e^{-i2z}) \int_0^1 \frac{dy}{y} (1 - e^{i2zy}) \right. \\ & \quad - e^{-i2z} \int_0^2 \frac{du}{u} (1 - e^{i2zu}) - 2i \sin 2z \log 2 \\ & \quad \left. + e^{i2z} \int_0^1 \frac{du}{u} (1 - e^{-i2zu}) \right], \end{aligned}$$

and the verification of Schelkunoff's relations (1.1) and (1.2) follows when (2.9) is used and the real and imaginary parts separated.

3. Auxiliary Summations.

The treatment of section 2 is incomplete in that the sum (2.4) has not been established. We shall now prove (2.4), regarding it as a special case ($\nu=0$, $b=-\mu$) of

$$\begin{aligned} & \sum_{m=0}^{\infty} \frac{(2\nu+1)_{2m+1} (b)_{2m+1}}{(2m+1)! (2+2\nu-b)_{2m+1}} \left[\frac{1}{\nu+2m+1} + \frac{1}{\nu+2m+2} \right] \\ & = \frac{\sqrt{\pi} \Gamma(2+2\nu-b)}{2^{2\nu+1} \Gamma(\nu+1/2) \Gamma(2+\nu-b)} \int_0^1 dx (1-x)^{\nu-1} (x^{-b}-1). \quad (3.1) \end{aligned}$$

The last integral may be expressed in terms of gamma functions.

In order to obtain (3.1) we start with the integral

$$\begin{aligned} & \int_0^1 dy y^{\nu-1} (1+y) [F(2\nu+1, b; 2+2\nu-b; y) \pm F(2\nu+1, b; 2+2\nu-b; -y)] \\ & = \sum_{n=0}^{\infty} \frac{(2\nu+1)_n (b)_n}{n! (2+2\nu-b)_n} (1 \pm (-)^n) \int_0^1 dy y^{\nu+n-1} (1+y) \\ & = \sum_{n=0}^{\infty} \frac{(2\nu+1)_n (b)_n}{n! (2+2\nu-b)_n} (1 \pm (-)^n) \left[\frac{1}{\nu+n} + \frac{1}{\nu+n+1} \right]. \quad (3.2) \end{aligned}$$

We may apply relations given by Kummer to the hypergeometric functions in the integrand and obtain

$$\begin{aligned} & \int_0^1 dy y^{\nu-1} (1+y)^{-2\nu} [(F(\nu+1/2, \nu+1; 2+2\nu-b; 4y(1+y)^{-2}) \\ & \quad \pm F(\nu+1/2, \nu+1-b; 2+2\nu-b; 4y(1+y)^{-2})]. \end{aligned}$$

Expanding the hypergeometric functions and using the results

$$\begin{aligned} & \int_0^1 \frac{dy y^{\nu+n-1}}{(1+y)^{2\nu+2\nu}} = \frac{1}{2} \frac{\Gamma(n+\nu) \Gamma(n+\nu)}{\Gamma(2n+2\nu)}, \\ & \frac{2^{2n-1} (\nu+1/2)_n \Gamma(n+\nu)}{\Gamma(2n+2\nu)} = \frac{\sqrt{\pi}}{2^{2\nu} \Gamma(\nu+1/2)}, \\ & F(a, b; c; 1) = \frac{\Gamma(c) \Gamma(c-a-b)}{\Gamma(c-a) \Gamma(c-b)}, \end{aligned}$$

$$\begin{aligned} \text{gives } \frac{\sqrt{\pi}}{2^{2\nu} \Gamma(\nu+1/2)} \sum_{n=0}^{\infty} \frac{\Gamma(n+\nu)}{n! (2+2\nu-b)_n} [(\nu+1)_n \pm (\nu+1-b)_n] \\ = \frac{\sqrt{\pi} \Gamma(2+2\nu-b)}{2^{2\nu} \Gamma(\nu+1/2) \Gamma(2+\nu-b)} \left[\frac{\Gamma(\nu) \Gamma(1-b)}{\Gamma(1+\nu-b)} \pm \frac{\Gamma(\nu) \Gamma(1)}{\Gamma(\nu+1)} \right]. \quad (3.3) \end{aligned}$$

Equating expressions (3.3) and (3.2), taking the lower (minus) sign, and using

$$\frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1} (1-x)^{b-1} dx, \quad (3.4)$$

gives (3.1).

Equating the expressions (3.2) and (3.3) really gives us expressions for the sums of two infinite series. They are

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{(2\nu+1)_n (b)_n}{n! (2+2\nu-b)_n} \left[\frac{1}{\nu+n} + \frac{1}{\nu+n+1} \right] \\ = \frac{\sqrt{\pi}}{2^{2\nu}} \frac{\Gamma(\nu) \Gamma(2+2\nu-b) \Gamma(1-b)}{\Gamma(\nu+1/2) \Gamma(2+\nu-b) \Gamma(1+\nu-b)}, \quad (3.5) \end{aligned}$$

$$\begin{aligned} \sum_{n=0}^{\infty} (-)^n \frac{(2\nu+1)_n (b)_n}{n! (2+2\nu-b)_n} \left[\frac{1}{\nu+n} + \frac{1}{\nu+n+1} \right] \\ = \frac{\Gamma(2+2\nu-b) \Gamma(\nu)}{\Gamma(2+\nu-b) \Gamma(2\nu+1)}. \quad (3.6) \end{aligned}$$

The second result is a special case of equation (4.5), a result given by Whipple.

When (3.5) and (3.6) are used to sum the second series in (2.2) and (2.3) the following equations are obtained.

$$\sum_{n=0}^{\infty} A_n = 2^{1-\alpha-\beta} \sum_{t=0}^{\infty} (-)^t \frac{z^{2t+\alpha+\beta} \left(\frac{\alpha+\beta}{2} \right)_{t+1}}{\Gamma(t+\alpha+1) \Gamma(t+\beta+1) \left(\frac{\alpha+\beta+1}{2} \right)_{t+1}}, \quad (3.7)$$

$$\begin{aligned} \sum_{n=0}^{\infty} (-)^n A_n &= 2^{2-\alpha-\beta} \sum_{t=0}^{\infty} (-)^t \frac{z^{2t+\alpha+\beta} \left(\frac{\alpha+\beta}{2} \right)_{t+1}}{\Gamma(t+\alpha+1) \Gamma(t+\beta+1) t! (2t+\alpha+\beta+1)} \\ A_n &= \frac{(\alpha+\beta-1)_{n+1}}{n!} \left(\frac{1}{\alpha+\beta-1+2n} + \frac{1}{\alpha+\beta+1+2n} \right) J_{n+\alpha}(z) J_{n+\beta}(z), \quad (3.8) \end{aligned}$$

$$\sum_{n=0}^{\infty} B_n = \frac{2^{2-\alpha-\beta} \Gamma\left(\frac{\alpha+\beta}{2}\right)}{\Gamma\left(\frac{\alpha-\beta}{2}\right)} \sum_{t=0}^{\infty} (-)^t \frac{z^{2t+\alpha+\beta} \left(\frac{\alpha+\beta}{2} \right)_{t+1}}{\Gamma(t+\alpha+1) \Gamma(t+\alpha+\beta+1) t! (2t+\alpha+\beta+1)}, \quad (3.9)$$

$$\begin{aligned} \sum_{n=0}^{\infty} (-)^n B_n &= \frac{\sqrt{\pi} \Gamma\left(\frac{\alpha-\beta+1}{2}\right)}{2^{\alpha-\beta-1} \Gamma\left(\frac{\alpha-\beta}{2}\right)} J_{\frac{\alpha+\beta+1}{2}}(z) J_{\frac{\alpha+\beta-1}{2}}(z) \\ B_n &= \frac{(\alpha-\beta-1)_{n+1}}{n!} \left(\frac{1}{\alpha-\beta-1+2n} + \frac{1}{\alpha-\beta+1+2n} \right) J_{n+\alpha}(z) J_{-n+\beta}(z). \quad (3.10) \end{aligned}$$

In (3.9) we must have $R(\beta) > -3/2$. For (3.10) to hold $(\alpha - \beta + 1)/2$ must not be zero or a negative integer, and if β is not an integer we must have $R(\beta) > -1$. The series on the left in (3.9) and (3.10) converge as $n \rightarrow \infty$ if β is an integer or if $R(\beta) > -3/2$ in (3.9) and $R(\beta) > -1$ in (3.10).

4. Additional Results*.

In this section we substitute values of a_n in (2.2) and (2.3) which enables us to perform one summation.

Setting $a_n = (\gamma)_n/n!$ in (2.2) and $a_n = (-)^n(\gamma)_n/n!$ in (2.3) gives

$$\sum_{n=0}^{\infty} \frac{(\gamma)_n}{n!} J_{n+\alpha}(z) J_{n+\beta}(z) = \sum_{t=0}^{\infty} \frac{(-)^t \Gamma(2t + \alpha + \beta - \gamma + 1)}{t! \Gamma(t + \alpha + 1) \Gamma(t + \beta + 1) \Gamma(t + \alpha + \beta - \gamma + 1)} \left(\frac{z}{2}\right)^{2t + \alpha + \beta}, \quad (4.1)$$

$$\sum_{n=0}^{\infty} (-)^n \frac{(\gamma)_n}{n!} J_{n+\alpha}(z) J_{-n+\beta}(z) = \sum_{t=0}^{\infty} \frac{(-)^t \Gamma(2t + \alpha + \beta - \gamma + 1) (z/2)^{2t + \alpha + \beta}}{t! \Gamma(t + \beta + 1) \Gamma(t + \alpha + \beta + 1) \Gamma(t + \alpha - \gamma + 1)}. \quad (4.2)$$

The series on the left in (4.2) converges if β is an integer, or if $R(\gamma) < R(\alpha + \beta + 1)$.

Another pair of series may be obtained from Dixon's sum †,

$${}_3F_2 \left[\begin{matrix} a, b, c \\ 1-a+c, 1-b+c \end{matrix} \right] = \frac{2^{-c} \sqrt{\pi} \Gamma(1-a+c) \Gamma(1-b+c) \Gamma(1-a-b+c/2)}{\Gamma\left(\frac{1+c}{2}\right) \Gamma\left(1-a+\frac{c}{2}\right) \Gamma\left(1-b+\frac{c}{2}\right) \Gamma(1-a-b+c)},$$

by assigning a and b appropriate values and using (2.2) and (2.3) :

$$\begin{aligned} & \sum_{n=0}^{\infty} \frac{(\alpha + \beta)_n (c)_n}{n! (1 + \alpha + \beta - c)_n} J_{n+\alpha}(z) J_{n+\beta}(z) \\ &= \sum_{t=0}^{\infty} (-)^t \frac{(\alpha + \beta + 1)_{2t} \left(1 + \frac{\alpha + \beta}{2} - c\right)_t \left(\frac{z}{2}\right)^{2t + \alpha + \beta}}{t! \Gamma(t + \alpha + 1) \Gamma(t + \beta + 1) (1 + \alpha + \beta - c)_t \left(1 + \frac{\alpha + \beta}{2}\right)_t}, \quad (4.3) \\ & \sum_{n=0}^{\infty} (-)^n \frac{(\alpha - \beta)_n (c)_n}{n! (1 + \alpha - \beta - c)_n} J_{n+\alpha}(z) J_{-n+\beta}(z) \end{aligned}$$

* A number of interesting results of a similar nature have been given by W. N. Bailey, "Some Series of Squares of Bessel Functions," Camb. Phil. Soc. Proc. xxvi. pp. 82-87 (1930).

† 'Generalized Hypergeometric Series,' W. N. Bailey (Camb. Math. Tract), p. 13.

$$\begin{aligned}
&= \frac{\Gamma\left(1+\frac{\alpha-\beta}{2}\right) \Gamma(1+\alpha-\beta-c)}{\Gamma(1+\alpha-\beta) \Gamma\left(1+\frac{\alpha-\beta}{2}-c\right)} \\
&\times \sum_{t=0}^{\infty} (-)^t \frac{\Gamma(\alpha+\beta+2t+1) \Gamma\left(1+\frac{\alpha+\beta}{2}-c+t\right) \left(\frac{z}{2}\right)^{2t+\alpha+\beta}}{t! \Gamma(\beta+t+1) \Gamma(\alpha+\beta+t+1) \Gamma\left(1+\frac{\alpha+\beta}{2}+t\right) \Gamma(\alpha-c+1+t)} . \\
&\quad \dots (4.4)
\end{aligned}$$

If, in (4.3), $\alpha+\beta-c$ is a negative integer we may change one of the quantities, say α , slightly and multiply both sides by the reciprocal of $\Gamma(1+\alpha+\beta-c)$. Letting α approach its original value shows that the terms for which n and t are less than $c-\alpha-\beta$ become zero and hence are to be omitted. The series on the left in (4.4) converges if $R(2c-2) < R(\alpha+\beta)$.

Finally, using Whipple's sum *

$${}_4F_3 \left[\begin{matrix} a, & 1+a/2, & d, & e, & -1 \\ a/2, & 1+a-d, & 1+a-e \end{matrix} \right] = \frac{\Gamma(1+a-d) \Gamma(1+a-e)}{\Gamma(1+a) \Gamma(1+a-d-e)}, \quad (4.5)$$

yields the results

$$\begin{aligned}
&\sum_{n=0}^{\infty} (-)^n (\alpha+\beta+2n) \frac{\Gamma(\alpha+\beta+n) (e)_n}{n! \Gamma(1+\alpha+\beta-e+n)} J_{n+\alpha}(z) J_{n+\beta}(z) \\
&= \sum_{t=0}^{\infty} (-)^t \frac{\Gamma(\alpha+\beta+2t+1) (z/2)^{2t+\alpha+\beta}}{t! \Gamma(\alpha+t+1) \Gamma(\beta+t+1) \Gamma(1+\alpha+\beta+t-e)}, \quad (4.6)
\end{aligned}$$

$$\begin{aligned}
&\sum_{n=0}^{\infty} (\alpha-\beta+2n) \frac{\Gamma(\alpha-\beta+n) (e)_n}{n! \Gamma(1+\alpha-\beta-e+n)} J_{n+\alpha}(z) J_{-n+\beta}(z) \\
&= \sum_{t=0}^{\infty} (-)^t \frac{\Gamma(2t+\alpha+\beta+1) (z/2)^{2t+\alpha+\beta}}{t! \Gamma(\beta+t+1) \Gamma(t+\alpha+\beta+1) \Gamma(t+\alpha-e+1)}. \quad (4.7)
\end{aligned}$$

The series on the left in (4.7) converges if $R(2e-2) > R(\alpha+\beta)$ or if β is an integer.

Some special cases of these results are

$$\begin{aligned}
&\sum_{n=0}^{\infty} J_{n+\alpha}(z) J_{-n+\alpha-1}(z) = \frac{1}{2} J_{2\alpha-1}(2z), \quad 0 < R(\alpha), \\
&\sum_{n=0}^{\infty} (-)^n \frac{(2\alpha-1)_n}{n!} J_{n+\alpha}(z) J_{n+\alpha-1}(z) = \frac{z^{\alpha-1}}{\Gamma(\alpha) 2^{2\alpha-1}} J_{\alpha}(2z), \\
&\sum_{n=0}^{\infty} \frac{J_n(z) J_{n+1}(z)}{2n+1} = \frac{1}{2z} \sin^2 z, \\
&\sum_{n=0}^{\infty} (-)^n \frac{(1/2-\alpha)_n}{(3/2+\alpha)_n} J_{n+\alpha}(z) J_{-n+\alpha-1}(z) = \frac{\Gamma(\alpha+3/2)}{2\Gamma(\alpha+1)} z^{-1/2} J_{2\alpha-1/2}(2z), \\
&\sum_{n=0}^{\infty} \frac{(\alpha-1/2)_n}{n!} J_{n+\alpha}(z) J_{n-1/2}(z) = \frac{1}{2^{\alpha-1/2} \sqrt{\pi z}} J_{\alpha}(2z), \quad 0 < R(\alpha),
\end{aligned}$$

* Bailey's Tract, *loc. cit.* p. 28.

$$\sum_{n=0}^{\infty} (\alpha+n) \frac{(2\alpha)_n}{n!} J_{n+\alpha}(z) J_{-n-\alpha}(z) = \left(\frac{z}{4}\right)^{\alpha} \frac{1}{\Gamma(\alpha)} J_{-\alpha}(2z), \quad R(\alpha) < \frac{1}{2},$$

$$\sum_{n=0}^{\infty} (-)^n (\alpha+n) \frac{(2\alpha)_n}{n!} J_{n+\alpha}^2(z) = \left(\frac{z}{4}\right)^{\alpha} \frac{1}{\Gamma(\alpha)} J_{\alpha}(2z),$$

$$\sum_{n=0}^{\infty} J_{n+\alpha}(z) J_{n+\beta}(z) = \frac{z}{2(\alpha-\beta)} [J_{\alpha-1}(z) J_{\beta}(z) - J_{\alpha}(z) J_{\beta-1}(z)].$$

The last result may be obtained more easily from the recurrence relations for Bessel functions. The first equation is a special case of a result given by D. G. Kendall*. The second and third equations from the last are special cases of Gegenbauer's addition theorem† and its modification (by Watson), respectively.

LXXXII. On the Temperature Coefficient of Air-cored Self-inductances.

By A. BLOCH, Dr. Ing., M.Sc.

(Communication from the Staff of the Research Laboratories of
The General Electric Company, Limited, England ‡.)

[Received May 23, 1944.]

SUMMARY.

Butterworth, in his classical paper on the high-frequency resistance of inductance coils, obtained the power losses in the coil wire by evaluating the real part of the complex form of Poynting's Integral taken over the surface of the wire. The imaginary part of the same integral was evaluated by T. J. Douma, who believed thus to have calculated the wattless power exchanged by the "internal" reactance of the coil. However, the penetration of the magnetic field into the conductor alters also the field surrounding the conductor, and it is the total change in the stored magnetic energy which is responsible for the "internal" reactance (*i. e.* the additional reactance caused by the fact that the resistance of the wire is not zero).

The present paper calculates the total effect by placing the surface of integration at infinite distance—or at least at such distance that the disturbance caused by the introduction of the conductor into the field has become negligibly small. The imaginary part of the integral tends then towards infinity. As the effect sought for is obtained from the

* Quart. J. Math. (Ox. S.) xiii. pp. 172–184.

† Watson, *loc. cit.* p. 363.

‡ Communicated by Dr. C. C. Paterson, F.R.S.

change caused in the value of this integral when the conductor is introduced into the field, care is required to define the boundary conditions at the surface of integration, which are kept constant when making this change. Hydrodynamic analogy, as used in "Part I." of this paper*, shows that at low frequencies the correction required amounts to 100 per cent. of the proximity contribution. At high frequencies the correction is zero.

The paper discusses also the variation of the coil inductance with change of the geometrical dimensions, including the case where the material of the wire has a coefficient of thermal expansion which differs from that of the coil former.

It is a well known fact that the inductance of a coil changes with temperature much more than one would expect from the thermal changes of the coil dimensions. The discrepancy may be of the order of several hundred per cent. The reason for this is that the change in temperature causes, amongst other things, a change in the resistivity of the coil wire (and of other conductors situated in the magnetic field of the coil) and influences thereby the extent to which the magnetic field penetrates into these conductors.

In a previous paper* the author has shown how to calculate this effect in a simple manner for the practically very important case in which the current flow is restricted to a thin skin at the surface of the conductor. The purpose of the present paper is to give a solution for the case where this assumption is no longer justified.

To calculate the reactance change associated with this varying penetration we require a knowledge of the magnetic field distribution around the coil wire†. Luckily, this is already known from the classical investigations of Butterworth‡ on the high-frequency resistance of such coils. It is useful for the following developments to recall the characteristic features and assumptions of the simplest type of Butterworth's solutions, which applies to coils in which the conductors are not too narrowly spaced. The field around each element of the coil wire is then considered to consist of two component fields, namely :

(a) a component H_0 which is caused by the current flowing in the wire element under consideration and which has field lines concentric with the wire element ;

* "On the Temperature Coefficient of Air-cored Self-inductances : the case of a 'thin' current-carrying layer," published in 'Wireless Engineer,' vol. xxi. pp. 359-367 (1944). This paper will hereafter be referred to as "Part I."

† We shall restrict ourselves here to the case of an unscreened coil wound with wire of circular cross-section.

‡ Phil Trans. Roy. Soc. vol. cxxii. (1922). See also a series of articles by the same author in the 'Wireless Engineer,' vol. iii. (1926) on "The Effective Resistance of Inductance Coils at Radio Frequencies."

(b) a field component H_1 which is caused by the current flowing in all the other elements of the coil. If we could remove from the coil the wire element under consideration (without interrupting the current flow through the rest of the coil) this field would fill the space of the removed element, and also the space adjacent to it, with a homogeneous field. The actual presence of the wire element causes a distortion of this field, and Butterworth assumes that this distortion is the same as if we had inserted the straightened coil wire into a homogeneous cross-field of infinite extension.

By evaluating Poynting's Integral for the energy flow across the surface of the coil wire Butterworth obtains the high-frequency resistance of the coil in the form

$$R=R_0 \left[1+F(z)+\gamma^2 \frac{d^2}{4} G(z) \right], \quad (1)$$

where

R_0 =d.c. resistance of the coil wire.

d =wire diameter.

$z=d\pi\sqrt{2f\rho}$ is a dimensionless constant*.

f =frequency.

ρ =resistivity in C.G.S. elect. magnetic units.

$F(z)$ and $G(z)$ are functions tabulated by Butterworth; $F(z)$ refers to the energy dissipation caused by the H_0 component of the field ("skin effect") and $G(z)$ to the dissipation caused by the H_1 component ("proximity effect"), γ^2 expresses the mean squared value of H_1 taken over the length of the coil wire. It can be calculated from

$$\gamma^2 = \frac{KN}{D}, \quad (2)$$

where K is a constant dependent on coil dimensions, N is the number of turns of the coil, and D is the outside diameter of the coil.

Tables for K were given by Butterworth, and graphs for it were included in Part I. of this paper. (Such graphs can also be found on page 81 of Terman's 'Radio Engineering Handbook'.)

Butterworth used the complex form of Poynting's Integral but evaluated only the real part of it, being only interested in the real power consumed within the conductor. The imaginary part of the same integral gives the wattless power associated with the field inside the conductor. This part was evaluated by T. J. Douma†, who thus assumed to have determined the internal reactance of the coil. This is true in an obviously

* It may be of interest that this constant can be interpreted as a measure of the ratio wire diameter/"skin-thickness" $z = \frac{d}{t} \frac{1}{\sqrt{2}}$, with $t = \frac{1}{2\pi} \sqrt{\frac{\rho}{f}}$ = equivalent thickness of conducting skin.

† "Internal Inductance of Coils and its Influence on the Temperature Coefficient of the Coil," 'Philips Transmitting News,' vol. v. (1938), nos. 1 and 2.

narrow sense, but not quite in the sense in which the concept of internal reactance is used in the present problem. Here we are interested in the increment of reactance caused by the penetration of the magnetic field into the conductor, and hence we have to take into account not only the magnetic energy stored within the bounds of the conductor, but also those changes in the external field brought about by this same penetration.

If this circumstance is taken into account, Douma's results have to be, in part, revised. Douma's results were expressed in the following form :

$$\omega L_4 = R_0 \left[F_1(z) + \gamma^2 \frac{d^2}{4} G_1(z) \right], \quad . \quad . \quad . \quad . \quad (3)$$

where $F_1(z)$ expresses the effect due to the H_0 -type of field and $G_1(z)$ the effect due to the H_1 -type of field.

Now if the H_0 -type of field penetrates into the conductor, the field-strength values outside the conductor remain unchanged ($H_0 = 2I/r$, where I = current in the conductor and r = distance from its axis); therefore the energy content of the field outside the conductor remains unchanged and Douma's method of calculation leads to a correct result. However, the same is not true for the H_1 -type of field; Douma's method gives us, therefore, in this case, only part of the effect which we wish to calculate. From the hydrodynamic analogy which we used in Part I. of this paper we can see at once that at low frequencies the correction required amounts to 100 per cent., for if the field penetrates the wire entirely the integral extended over the surface of the wire can only give another expression for the energy stored in the volume of the wire; we know, however, that the actual energy difference in this case amounts to just twice this amount. On the other hand, if the concept of the equivalent magnetic skin which we used for the high-frequency case gave us also the correct reaction on the external field, the correction at high frequencies will be zero.

In the following investigation the total effect will be calculated, assuming, as before, that the external field extends as far as infinity, *i. e.* we shall evaluate Poynting's integral over a cylindrical surface parallel to the axis of the wire and of infinite radius. On account of the infinite amount of magnetic energy stored in this volume, the imaginary part of the integral will in this case, of course, be infinite. We are interested here in the difference between the infinite values obtaining before and after the introduction of the conductor into the field. Just as in the analogous hydrodynamical case, care is required to define clearly the conditions at the infinite boundary which are kept constant when making this change*.

* This procedure is quite in keeping with the original assumption underlying Butterworth's solution, though it may appear at first sight to be less convincing. However, it can easily be seen that the following developments apply also to the case of a surface of integration at finite distance, provided this distance is so large that the field disturbance caused by the introduction of a conductor has become small of first order.

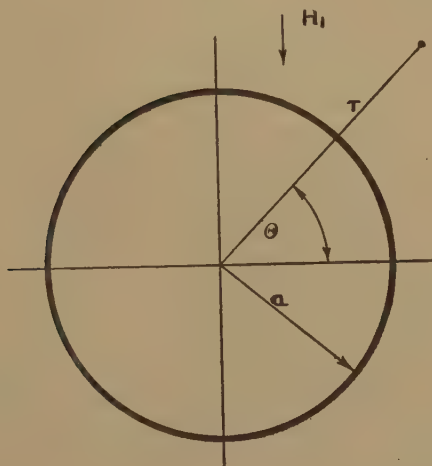
Using the co-ordinate system shown in the figure, we can specialize Butterworth's solution for the present case as follows :

For $r > a$ we have

$$\hat{P}_1 = \left[1 + \left(\frac{a}{r} \right)^2 \hat{\chi}_1 \right] \hat{H}_1 \sin \Theta, \quad (4)$$

$$\hat{Q}_1 = \left[1 - \left(\frac{a}{r} \right)^2 \hat{\chi}_1 \right] \hat{H}_1 \cos \Theta, \quad (5)$$

$$E_1 = j\omega r \left[1 + \left(\frac{a}{r} \right)^2 \hat{\chi}_1 \right] \hat{H}_1 \cos \Theta, \quad (6)$$



Here, in complex notation (suppressing as usual the factor $e^{j\omega t}$),

\hat{P}_1 denotes the radial component of the magnetic field;

\hat{Q}_1 denotes the tangential component of the magnetic field;

\hat{E}_1 denotes the electric field strength in a direction parallel to the axis of the cylinder;

a is the radius of the cylinder:

$$\chi_1 = \frac{J_2(z\sqrt{-j})}{J_0(z\sqrt{-j})}. \quad (7)$$

One verifies very easily—at least for \hat{P}_1 and \hat{Q}_1 —that a homogeneous field is described by equations of the same form, but with the terms in χ_1 missing. These terms express, then, the action on the external field by the eddy currents in the conductor, and we notice that at infinite distance these contributions become infinitely small, of second order. At infinite distance the field strength approaches, therefore, the value of the undisturbed field—as one would expect from physical considerations.

The value of Poynting's Integral is

$$\Pi = \frac{1}{8\pi} \int_0^{2\pi} \hat{E}_1 \check{Q}_1 r d\Theta, \quad (8)$$

where \check{Q} denotes the complex conjugate of \hat{Q} .

Inserting values, we get

$$\Pi = \frac{1}{8\pi} \int_0^{2\pi} j\omega r \hat{H}_1 \left[1 + \left(\frac{a}{r} \right)^2 \hat{\chi}_1 \right] \cos \Theta \cdot \check{H}_1 \left[1 - \left(\frac{a}{r} \right)^2 \check{\chi}_1 \right] \cos \Theta \cdot r d\Theta. \quad (9)$$

If we put $r=a$, we obtain Douma's result, and it is worth while for comparison to derive it. We get

$$\Pi_D = \frac{|\hat{H}_1|^2 \cdot j\omega a^2}{8} [1 - \hat{\chi}_1 \check{\chi}_1] + (\hat{\chi}_1 - \check{\chi}_1) = \Delta W_{\text{mean}} + j2\omega \cdot \Delta T_{\text{mean}}. \quad (10)$$

with

$$\hat{\chi}_1 = -\Phi_1 - j\Psi_1, \quad (11)$$

and splitting into real and imaginary parts we arrive at

$$\Delta W_{\text{mean}} = |\hat{H}_1|^2 \cdot \frac{\omega a^2}{4} \cdot \Psi_1, \quad (12)$$

$$\Delta T_{\text{mean}} = |\hat{H}_1|^2 \frac{a^2}{16} (1 - \Phi_1^2 - \Psi_1^2). \quad (13)$$

$$\begin{array}{ll} \text{For } z=0 \text{ we have} & \Phi_1=0, \quad \Psi_1=0, \\ \text{and for } z=\infty \text{ we have} & \Phi_1=1, \quad \Psi_1=0. \end{array} \quad (14)$$

Thus the value of ΔT_{mean} (the mean energy stored in the volume of the wire) ranges in this case for a variation of z from zero (complete penetration of the field) to infinity (complete exclusion of the field) just over $|\hat{H}_1|^2 \cdot a^2/16$ (the mean energy content of the wire volume if filled by a homogeneous field \hat{H}_1).

Now let the radius of the cylindrical surface of integration grow towards infinity.

We have

$$\Pi = \frac{|\hat{H}_1|^2}{8} \cdot j\omega \left[1 - \left(\frac{a}{r} \right)^4 \hat{\chi}_1 \check{\chi}_1 + \left(\frac{a}{r} \right)^2 (\hat{\chi}_1 - \check{\chi}_1) \right] r^2, \quad . . . (15)$$

which for $r=R \rightarrow \infty$ gives

$$\Pi = \frac{|\hat{H}_1|^2}{8} j\omega R^2 + \frac{|\hat{H}_1|^2}{8} j\omega a^2 (\hat{\chi}_1 - \check{\chi}_1). \quad (16)$$

The second term of this expression, which gives the real power consumption, is of course the same as found above from eqn. (10). The first term, which tends towards infinity, has to be compared with the value which would have been obtained in the absence of the conductor,

$$\Pi_0 = \frac{|\hat{H}_{10}|^2}{8} \cdot j\omega R^2, \quad (17)$$

where \hat{H}_{10} denotes the value of the undisturbed field,

On account of the factor R^2 , the infinitely small difference between H_1 and \hat{H}_{10} cannot be neglected, and it is at this point that a careful definition of the conditions at the cylindrical boundary is required.

Let us assume here that the field is produced by a winding of suitable "density" placed upon the infinite cylindrical shell (wires running parallel to its axis), and let us keep the current through this winding constant while we introduce our cylindrical conductor. This amounts to stipulating that we have kept constant at infinity the value of the tangential component of the field, *i.e.*

$$\hat{Q}_1 = \hat{Q}_{10}, \quad \dots \quad (18)$$

and from this follows

$$\hat{H}_1 \left[1 - \left(\frac{a}{R} \right)^2 \hat{\chi}_1 \right] \hat{H}_{10} \quad \dots \quad (19)$$

or

$$\hat{H}_1 = \frac{\hat{H}_{10}}{1 - \left(\frac{a}{R} \right)^2 \hat{\chi}_1} = \hat{H}_{10} \left[1 + \left(\frac{a}{R} \right)^2 \hat{\chi}_1 \right] \quad \dots \quad (20)$$

Thus we get now

$$\left. \begin{aligned} \Delta T_{\text{mean}} &= \frac{R^2}{16} [|\hat{H}_1|^2 - |H_{10}|^2] \\ &= \frac{R^2}{16} [\hat{H}_1 \cdot \check{H}_1 - \hat{H}_{10} \cdot \check{H}_{10}] \\ &= \frac{R^2}{16} \hat{H}_{10} \cdot \check{H}_{10} \cdot \frac{a^2}{R^2} (\hat{\chi}_1 + \check{\chi}_1) = -|\hat{H}_{10}|^2 \cdot \frac{a^2}{8} \cdot \Phi_1 \end{aligned} \right\} \quad \dots \quad (21)$$

We notice that for the same range of variation of z as above we get now a variation of magnetic energy which is twice as big—just as demanded by the hydrodynamic analogy used in Part I.

The energy change following the introduction of the cylindrical conductor (with constant excitation[†] of the field) is a decrease in energy; the maximum decrease in energy being obtained for $z = \infty$ *. In our definition of internal inductance we take this case as the datum, and we get, therefore, as mean energy increase due to the finite resistivity of the conductor

$$\Delta \bar{T}_{\text{int}} = |\hat{H}_{10}|^2 \cdot \frac{a^2}{8} (1 - \Phi_1) \quad \dots \quad (22)$$

\hat{H}_{10} has not the same value for all elements of the coil wire. If we

* If we had stipulated that the flux through the exciting coil is kept constant (*i.e.* if we had connected this coil to a source of alternating E.M.F. of constant amplitude) we would have had $\hat{P}_1 = \hat{P}_{10}$, and had found then an increase in energy of the same amount.

denote by \hat{H}_{1m} its mean squared value (taken over the length of the wire) we arrive at an incremental reactance

$$X_i = 4\omega \frac{\Sigma \Delta \bar{T}_{int}}{|\hat{I}|^2} = 4\omega \frac{|\hat{H}_{1m}|^2}{|\hat{I}|^2} \frac{la^2}{8} (1 - \Phi_1), \quad (23)$$

where \hat{I} denotes the current responsible for \hat{H}_{1m} .

With
$$\hat{H}_{1m} = \frac{KN}{D} \cdot \hat{I} = \gamma \cdot \hat{I}, \quad (24)$$

$$z^2 = \frac{4\pi\omega a^2}{\rho}, \quad (25)$$

and
$$R_0 = \frac{l \cdot \rho}{a^2 \pi}, \quad (26)$$

eqn. (23) can be rewritten as

$$X_i = R_0 \cdot \frac{d^2}{4} \cdot \gamma^2 \cdot G_{1B}, \quad (27)$$

where
$$G_{1B} = \frac{z^2}{8} (1 - \Phi_1), \quad (28)$$

in contrast to Douma's function

$$G_1 = \frac{z^2}{8} \cdot \frac{(1 - \Phi_1^2 - \Psi_1)}{2}. \quad (29)$$

For the evaluation of $\Phi_1(z)$ there are a number of possibilities, some of which are listed in the appendix together with formulæ for $G'_{1B}(z)$. For the convenience of the reader, formulæ for $F_1(z)$ and $zF_1'(z)$ are also given.

From the series development given in the Appendix follows a verification of the result derived in Part I. for the H_1 -field. For large values of z

$$G_{1B}(z) = \frac{z}{4\sqrt{2}} \quad (30)$$

and
$$1 - \Phi_1(z) = \frac{2}{z\sqrt{2}} \quad (31)$$

Therefore, according to eqn. (22). the "internal" energy per unit length of conductor is

$$\Delta \bar{T}_{int} = |\hat{H}_{10}|^2 \cdot \frac{a^2}{8} \cdot \frac{\sqrt{2}}{z} \quad (32)$$

Making use of the explanation given in the footnote on p. 695, this can be transformed into

$$\Delta \bar{T}_{int} = \frac{|\hat{H}_{10}|^2 \cdot a^2}{8} \cdot \frac{2t\sqrt{2}}{\sqrt{2} \cdot 2a} = |\hat{H}_{10}|^2 \cdot \frac{a \cdot t_m}{4} \quad (33)$$

This, however, is the result which we derive under the assumption of the equivalent magnetic skin thickness $t_m = t/2$.

We shall now discuss the temperature coefficient of the coil inductance in somewhat more detail than was done in Douma's paper.

If we assume the geometrical dimensions to remain constant, and only the resistivity to vary, we get from eqn. (3) *

$$\left. \begin{aligned} \omega \left(\frac{dL_i}{dT} \right)_\rho &= \frac{dR_0}{dT} \left[F_1 + \gamma^2 \frac{d^2}{4} G_{1B} \right] + R_0 \left[F'_1 + \gamma^2 \frac{d^2}{4} G'_{1B} \right] \frac{dz}{dT} \\ &= \frac{1}{R_0} \frac{dR_0}{dT} \omega L_i + R_0 \left[z F'_1 + \gamma^2 \frac{d^2}{4} z G'_{1B} \right] \frac{1}{z} \frac{dz}{dT} \\ &= \alpha \omega L_i - R_0 \left[z F'_1 + \gamma^2 \frac{d^2}{4} z G'_{1B} \right] \frac{\alpha}{2}, \end{aligned} \right\} \quad (34)$$

where the dashes denote differentiation with regard to z and

$$\alpha = \frac{1}{\rho} \frac{d\rho}{dT}. \quad (35)$$

For very large values of z , F_1 and G_{1B} are simply proportional to z ; thus $zF'_1 = F_1$ and $zG'_{1B} = G_{1B}$, and we get

$$\frac{1}{L_i} \left(\frac{dL_i}{dT} \right)_\rho = \alpha - \frac{\alpha}{2} = \frac{\alpha}{2}, \quad (36)$$

in confirmation of the results of Part I.

On the other hand, for very small values of z , G_{1B} and F_1 are proportional to z^2 , i. e., $zF'_1 = 2F_1$ and $zG'_{1B} = 2G_{1B}$. We therefore get

$$\frac{1}{L_i} \left(\frac{dL_i}{dT} \right)_\rho = 0. \quad (37)$$

In an analogous manner to that used for the derivation of eqn. (34) we get for a mere change in geometrical dimensions (coefficient of geometrical expansion β_1)

$$\omega \left(\frac{dL_i}{dT} \right)_{\beta_1} = -\beta_1 \omega L_i + R_0 \left[z F'_1 + \gamma^2 \frac{d^2}{4} z G'_{1B} \right] \beta_1. \quad (38)$$

We see that for large values of z this variation goes to zero, as already shown in section (3) of Part I. of this paper.

This variation of inductance is, of course, to be taken together with the variation of the external inductance L_0 , usually so much larger that there is no point in calculating the value of (38): we have

$$\omega \left(\frac{dL_0}{dT} \right)_{\beta_1} = \omega L_0 \beta_1. \quad (39)$$

Finally, let us consider the case (also already considered in Part I.) that the uniform geometrical expansion is followed by an expansion of

* We enclose ordinary differentiation signs in brackets, indicating at the base of the bracket the magnitude which is supposed to vary.

the wire diameter only (coefficient of expansion $\beta_2 - \beta_1 = \beta_3$). We have then

$$\omega \left(\frac{dL_i}{dT} \right)_{\beta_3} = -2\beta_3 \omega L_i + R_0 \left[zF'_1 + \gamma^2 \frac{d^2}{4} zG'_{1B} \right] \beta_3 + R_0 \gamma \frac{d^2}{4} G_{1B} 2\beta_3. \quad (40)$$

This last expression gives the inductance change against the zero resistivity case as datum value. To get the total change we have to add the change of this value, *i. e.* the inductance change caused by the expansion of the wire, if this wire is of zero resistivity.

This latter change, however, can be simply obtained from the expression for the energy stored in the annular volume, by which the external space of the coil is diminished. We get

$$\Delta(\omega L_0) = \left. \begin{aligned} & \frac{-2\omega}{8\pi|\hat{I}|^2} [|\hat{H}_0|^2 + 2|\hat{H}_{1m}|^2] 2a^2 l \beta_3 \pi \Delta T \\ & = \frac{-2\omega}{8|\hat{I}|^2} \left[\left(\frac{2\hat{I}}{a} \right)^2 + 2 \left(\frac{KN\hat{I}}{D} \right)^2 \right] 2a^2 l \beta_3 \Delta T. \end{aligned} \right\} \dots (41)$$

With the abbreviation (already introduced in Part I.)

$$\eta = 1 + \frac{1}{2} \left(\frac{KNd}{D} \right)^2, \quad \dots (42)$$

this leads finally to

$$\omega \left(\frac{dL_0}{dT} \right)_{\beta_3} = -2\omega l \eta \beta_3 \text{ (C.G.S. units).} \quad \dots (43)$$

APPENDIX.

$$F_1 = \frac{z^2}{8} (1 - \Phi_2), \quad \dots (A1)$$

where
$$-\Phi_2 - i\Psi_2 = \frac{J_3(z\sqrt{-j})}{J_1(z\sqrt{-j})} \dots (A2)$$

$$G_{1B} = \frac{z^2}{8} (1 - \Phi_1), \quad \dots (A3)$$

where
$$-\Phi_1 - i\Psi_1 = \frac{J_2(z\sqrt{-j})}{J_0(z\sqrt{-j})} \dots (A4)$$

For the evaluation of Φ_1 and Φ_2 in the ranges $z < 1$ and $z > 5$ Butterworth developed convenient series * which form the basis of the following formulæ :

For $z < 1$:

$$F_1 = \frac{z^2}{8} - \frac{z^6}{3072} + \frac{13z^{10}}{8847360} + \dots \dots (A5)$$

$$zF'_1 = \frac{z^2}{4} - \frac{z^6}{512} + \frac{13z^{10}}{884736} + \dots \dots (A6)$$

* Proc. Phys. Soc. vol. xxv. p. 294 (1913).

$$G_{1B} = \frac{z^2}{8} - \frac{z^6}{384} + \frac{19z^{10}}{983040} + \dots \quad (A7)$$

$$zG'_{1B} = \frac{z^2}{4} - \frac{z^6}{64} + \frac{19z^{10}}{98304} + \dots \quad (A8)$$

For $z > 5$:

$$F_1 = \frac{z}{2\sqrt{2}} - \frac{3}{16\sqrt{2}} \cdot \frac{1}{z} - \frac{3}{16} \cdot \frac{1}{z^2} - \frac{63}{256\sqrt{2}} \cdot \frac{1}{z^3}, \quad (A9)$$

$$zF'_1 = \frac{z}{2\sqrt{2}} + \frac{3}{16\sqrt{2}} \cdot \frac{1}{z} + \frac{3}{8} \cdot \frac{1}{z^2} + \frac{189}{256\sqrt{2}} \cdot \frac{1}{z^3}, \quad (A10)$$

$$G_{1B} = \frac{z}{4\sqrt{2}} + \frac{1}{32\sqrt{2}} \cdot \frac{1}{z} + \frac{1}{32} \cdot \frac{1}{z^2} + \frac{25}{128\sqrt{2}} \cdot \frac{1}{z^3}, \quad (A11)$$

$$zG'_{1B} = \frac{z}{4\sqrt{2}} - \frac{1}{32\sqrt{2}} \cdot \frac{1}{z} - \frac{1}{16} \cdot \frac{1}{z^2} - \frac{75}{128\sqrt{2}} \cdot \frac{1}{z^3}. \quad (A12)$$

$\frac{8}{z^2} \cdot F_1$ is tabulated in 'Bulletin Bureau of Standards,' vol viii. pp. 226-227 (1912) (denoted there by $\frac{4}{X} \cdot \frac{Z}{Y}$).

Further tabulation exists* for the functions $X(z)$, $V(z)$, $Z(z)$, $W(z)$ and for their combinations W/X and Z/V , which are related to the desired values Φ_1 and Φ_2 as follows :

$$\Phi_1 = 1 - \frac{2}{z} \cdot \frac{W(z)}{X(z)} \quad (A13)$$

$$\Phi_2 = 1 - \frac{4}{z} \cdot \frac{Z(z)}{V(z)}. \quad (A14)$$

Douma derives from the recurrence relation between Bessel Functions

$$1 - \Phi_1 - j\Psi_1 = \frac{2}{(z\sqrt{-j})} \cdot \frac{J_1(z\sqrt{-j})}{J_0(z\sqrt{-j})}, \quad (A15)$$

$$1 - \Phi_2 - j\left(\Psi_2 + \frac{8}{z^2}\right) = \frac{-4}{(z\sqrt{-j})} \cdot \frac{J_0(z\sqrt{-j})}{J_1(z\sqrt{-j})}. \quad (A16)$$

These can be conveniently evaluated when the values of $J_1(z\sqrt{-j})$ and $J_0(z\sqrt{-j})$ are available in polar form. The best tables for this purpose are given in the "Report of the British Association for the Advancement of Science" (Report of the 91st meeting, 1923). The tables give directly to 6 decimals the absolute magnitude and the real and imaginary part of $\frac{z}{2} \cdot \sqrt{-j} \cdot \frac{J_0}{J_1}$ under the headings M , $M \cos \beta$ and $M \sin \beta$.

* Savidge, Phil. Mag. vol. xix. p. 49 (1910). [5 decimals, integral values of the argument 0-30.]

TABLES of F_1 and G_{1B} .

Z.	F_1 .	G_{1B} .	Z.	F_1 .	G_{1B} .
0.0	0.00 000	0.00 000	5.0	1.73 740	0.88 767
0.1	0.00 125	0.00 125	5.1	1.77 296	0.90 545
0.2	0.00 500	0.00 500	5.2	1.80 849	0.92 324
0.3	0.01 125	0.01 125	5.3	1.84 403	0.94 103
0.4	0.02 000	0.01 999	5.4	1.87 958	0.95 882
0.5	0.03 124	0.03 121	5.5	1.91 515	0.97 659
0.6	0.04 498	0.04 488	5.6	1.95 074	0.99 435
0.7	0.06 121	0.06 095	5.7	1.98 637	1.01 210
0.8	0.07 991	0.07 933	5.8	2.02 202	1.02 982
0.9	0.10 108	0.09 989	5.9	2.05 771	1.04 753
1.0	0.12 468	0.12 247	6.0	2.09 343	1.06 522
1.1	0.15 068	0.14 683	6.1	2.12 916	1.08 289
1.2	0.17 904	0.17 267	6.2	2.16 492	1.10 055
1.3	0.20 970	0.19 966	6.3	2.20 070	1.11 819
1.4	0.24 259	0.22 740	6.4	2.23 649	1.13 582
1.5	0.27 762	0.25 546	6.5	2.27 229	1.15 345
1.6	0.31 470	0.28 342	6.6	2.30 809	1.17 106
1.7	0.35 368	0.31 086	6.7	2.34 390	1.18 867
1.8	0.39 443	0.33 744	6.8	2.37 970	1.20 627
1.9	0.43 679	0.36 285	6.9	2.41 550	1.22 388
2.0	0.48 057	0.38 689	7.0	2.45 129	1.24 148
2.1	0.52 558	0.40 945	7.1	2.48 708	1.25 907
2.2	0.57 161	0.43 052	7.2	2.52 285	1.27 667
2.3	0.61 845	0.45 015	7.3	2.55 861	1.29 427
2.4	0.66 587	0.46 847	7.4	2.59 436	1.31 188
2.5	0.71 365	0.48 565	7.5	2.63 010	1.32 948
2.6	0.76 156	0.50 189	7.6	2.66 582	1.34 709
2.7	0.80 942	0.51 739	7.7	2.70 153	1.36 470
2.8	0.85 702	0.53 235	7.8	2.73 723	1.38 231
2.9	0.90 420	0.54 696	7.9	2.77 292	1.39 992
3.0	0.95 081	0.56 139	8.0	2.80 860	1.41 754
3.1	0.99 674	0.57 576	8.1	2.84 427	1.43 515
3.2	1.04 189	0.59 019	8.2	2.87 993	1.45 277
3.3	1.08 619	0.60 477	8.3	2.91 558	1.47 039
3.4	1.12 963	0.61 956	8.4	2.95 123	1.48 801
3.5	1.17 217	0.63 461	8.5	2.98 687	1.50 563
3.6	1.21 385	0.64 993	8.6	3.02 252	1.52 325
3.7	1.25 468	0.66 555	8.7	3.05 816	1.54 087
3.8	1.29 471	0.68 145	8.8	3.09 381	1.55 848
3.9	1.33 401	0.69 763	8.9	3.12 946	1.57 609
4.0	1.37 265	0.71 408	9.0	3.16 513	1.59 369
4.1	1.41 068	0.73 076	9.1	3.20 081	1.61 128
4.2	1.44 819	0.74 767	9.2	3.23 650	1.62 886
4.3	1.48 525	0.76 476	9.3	3.27 221	1.64 643
4.4	1.52 194	0.78 202	9.4	3.30 795	1.66 399
4.5	1.55 831	0.79 942	9.5	3.34 372	1.68 153
4.6	1.59 443	0.81 692	9.6	3.37 952	1.69 906
4.7	1.63 035	0.83 453	9.7	3.41 536	1.71 656
4.8	1.66 613	0.85 220	9.8	3.45 124	1.73 404
4.9	1.70 180	0.86 992	9.9	3.48 716	1.75 150
			10.0	3.52 312	1.76 893

From these tables we get directly

$$F_1 M \sin \beta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (A17)$$

for $\frac{z^2}{8} \text{Re} \frac{-4}{z\sqrt{-j}} \cdot \frac{J_0}{J_1} = \text{Re} \frac{-z}{2\sqrt{-j}} \cdot \frac{J_0}{J_1} = \mathcal{I}m \frac{z\sqrt{-j}}{2} \frac{J_0}{J_1}$. (A18)

G_{1B} can also be found without recourse to other tables :

$$\text{If } \operatorname{Re} \frac{z\sqrt{-j}}{2} \frac{J_0}{J_1} = M \cos \beta, \quad \dots \dots \dots (\text{A } 19)$$

then
$$\operatorname{Re} \frac{2}{z\sqrt{-j}} \frac{J_1}{J_0} = \frac{1}{M} \cos \beta \dots \dots \dots (\text{A } 20)$$

$$\text{and } G_{1B} = \frac{z^2}{8} \operatorname{Re} \frac{2}{z\sqrt{-j}} \cdot \frac{J_1}{J_0} = \frac{z^2}{8} \cdot \frac{\cos \beta}{M} = \frac{z^2}{8} \cdot \frac{M \cos \beta}{M^2} \quad (\text{A } 21)$$

The Table given on opposite page is based on these tables.

Example.

As an example of the application of the foregoing developments we shall calculate the temperature coefficient of inductance for a coil which had been measured at the Laboratories (see also the example and the remarks made in Part I. of this paper).

This coil had 82 turns of 32 S.W.G. copper wire ($d=0.193$ mm.) which were wound under tension on a ceramic former of 2.86 cm. diameter and 4.1 cm. length. The inductance of the coil was $105 \mu\text{H}$.

At 1.6 Mc./sec. (the frequency at which the measurements were made) the equivalent skin thickness for copper is 0.052 mm., *i. e.* we have a value of $z=d/t\sqrt{2}=2.6$. The graphs give $K=3.5$.

From this follows

$$\omega L_i = R_0 \left[F_1 + \frac{(KNd)^2}{2D} G_{1B} \right] = 4.37 [0.76 + 0.94 \times 0.50] = 5.35 \Omega, \quad (A 19)$$

i. e. the internal reactance of the coil amounts to 4.14×10^{-3} parts of the total reactance of the coil.

From eqn. (34) the temperature coefficient of internal inductance is

$$\begin{aligned} \frac{1}{\omega L_i} \cdot \frac{d(\omega L_i)}{dT} &= \alpha - \frac{R_0}{\omega L_i} \left[zF'_1 + \left(\frac{KNd}{2D} \right)^2 zG'_{1B} \right] \frac{\alpha}{2} \\ &= 4200 \times 10^{-6} - \frac{4 \cdot 37}{5 \cdot 35} [1 \cdot 23 + 0 \cdot 94 \times 0 \cdot 40] \times 2100 \times 10^{-6} = 1440 \times 10^{-6}. \quad (\text{A } 20) \end{aligned}$$

This means that the variation of the internal inductance alone will give to the coil as a whole a temperature coefficient of inductance equal to $1440 \times 4 \cdot 14 \times 10^{-3} \times 10^{-6} = 6 \times 10^{-6}$ parts per degree C. If we add to this value 8×10^{-6} , on account of the linear expansion of the coil, we arrive at a total temperature coefficient equal to 14×10^{-6} parts per degree C. To this we may have to add, presumably, 1 or 2×10^{-6} parts on account of the temperature change of the self-capacitance of the coil.

The measured value was 15.9×10^{-6} parts per degree C.

LXXXIII. *On the Theory of Seismometers.*

By C. G. PENDSE, M.A., Ph.D.*

[Received August 25, 1943.]

Introduction.

IN the theory of seismometers developed in seismological writings, full consideration does not appear to have been given to the fact that the motion recorded by a seismometer is relative to a frame determined by the instrument, whereas the system of forces in question refers to axes "fixed in the earth" relative to which the frame determined by the instrument also moves. In this paper it is proposed to obtain the exact equation of motion of a seismometer and to investigate it in an important particular case. The present work will apply mainly to horizontal seismometers. In § 1 the exact equation is obtained in its most general form and is simplified without any loss of generality; an important particular case is taken and the corresponding equation (D) is derived. In § 2 and § 3 this equation (D) is solved as far as the first order of the acceleration terms. In § 4, the results obtained are discussed in relation to seismometers and seismograms. Finally in § 5 another form for the ground motion is briefly considered.

The new features shown by the investigation include the existence of resonance periods other than the free period of the instrument.

§ 1. *The General Equation of Motion of a Seismometer.*

Usually the main principle in a seismometer is the motion of a rigid body about an axis. But the frame of the instrument will move relative to "axes fixed in the earth." Hence, dynamically, the problem is one of the rotation of a rigid body about an axis, which has an accelerated motion relative to the earth under the action of gravity and other forces.

Let m be the mass of the body, G its centre of mass, O the foot of the perpendicular from G to the axis about which the body can turn, and I the moment of inertia of the body about the axis. Let $OG=l$.

Let \dagger OX, OY, OZ be axes fixed in the body, OZ being along the axis

* Communicated by the Author.

† The axes are Cartesian rectangular. The expression "axes fixed in the earth" refers to the frame to which the gravity field refers. This explanation (or definition) is necessary because in an earthquake the different portions of the earth move relative to each other. The directions of Ox, Oy, Oz are fixed relative to these axes "fixed in the earth."

of rotation, OX being along OG and OY being such that the frame OXYZ is right-handed. Let $Oxyz$ be another right-handed frame through O such that Oz is along the upwards vertical, Ox, Oy being in the horizontal plane such that their directions are fixed relative to the earth. Let ϕ be the angle between the planes Ozx and OZz, θ the angle ZOz and ψ the angle between the planes ZOz and ZOX; so that ϕ, θ, ψ are the Eulerian co-ordinates of the frame OXYZ, *i. e.* the body, relative to the frame $Oxyz$. Let $\mathbf{V}_x, \mathbf{V}_y, \mathbf{V}_z, \mathbf{V}_X, \mathbf{V}_Y, \mathbf{V}_Z$ be unit-vectors along the directions $\vec{Ox}, \vec{Oy}, \vec{Oz}, \vec{OX}, \vec{OY}, \vec{OZ}$ respectively, and let g denote the acceleration due to gravity. Then we have the following relations:—

$$\mathbf{V}_X = (\cos \psi \cos \theta \cos \phi - \sin \phi \sin \psi) \mathbf{V}_x + (\cos \theta \sin \phi \cos \psi + \cos \phi \sin \psi) \mathbf{V}_y - \cos \psi \sin \theta \mathbf{V}_z,$$

$$\mathbf{V}_Z = \sin \theta \cos \phi \mathbf{V}_x + \sin \theta \sin \phi \mathbf{V}_y + \cos \theta \mathbf{V}_z.$$

It will be convenient to find out here the expression for the moment of a vector localized at G about \vec{OZ} in terms of its resolutes along $\vec{Ox}, \vec{Oy}, \vec{Oz}$. Let $\mathbf{F} = F_x \mathbf{V}_x + F_y \mathbf{V}_y + F_z \mathbf{V}_z$.

The moment of \mathbf{F} localised at G about \vec{OZ}

$$= (\mathbf{V}_Z, [\vec{OG}, \mathbf{F}])$$

$$= l \begin{vmatrix} \sin \theta \cos \phi, & \sin \theta \sin \phi, & \cos \theta \\ \cos \theta \cos \phi \cos \psi - \sin \theta \sin \psi, & \cos \theta \sin \phi \cos \psi + \cos \phi \sin \psi, & -\cos \psi \sin \theta \\ F_x, & F_y, & F_z \end{vmatrix}$$

We consider the motion of the system relative to the frame OXYZ when O has an acceleration $\mathbf{a} = a_x \mathbf{V}_x + a_y \mathbf{V}_y + a_z \mathbf{V}_z$, not necessarily constant, relative to axes fixed in the earth, and when the orientation of OXYZ remains invariable relative to axes fixed in the earth*. By the principles of dynamics, the system of forces on the body, for motion relative to OXYZ, is composed of (1) the reaction of the axis, (2) the forces due to gravity, equivalent to a force $-mg$ at G parallel to \vec{Oz} , (3) the force $-ma$ at G and (4) other forces, if any. Let M be the moment of the forces under (4) about the axis \vec{OZ} . It is presumed that the reaction of the axis has no moment about \vec{OZ} ; in case the forces of reaction do not reduce to a single force intersecting OZ, the resulting

* In the most general case the frame OXYZ will have rotational motion relative to $Oxyz$ in addition to O having acceleration relative to axes fixed in the earth. This case is not considered here. It may be mentioned that the problem will become much more complicated than now as the axes OX, OY, OZ may not be the principal axes of inertia.

moment can be included under (4). The equation of motion of the body about the axis OZ is then

$$I\ddot{\psi} = M - ml \begin{vmatrix} \sin \theta \cos \phi, & \sin \theta \sin \phi, \\ \cos \theta \cos \phi \cos \psi - \sin \theta \sin \psi, & \cos \theta \sin \phi \cos \psi + \cos \phi \sin \psi, \\ a_x, & a_y, \\ & \cos \theta \\ & -\cos \psi \sin \theta \\ & a_z + g \end{vmatrix}. \quad (A)$$

Without any loss of generality of (A) we can put $\phi=0$, since ϕ remains constant. Then (A) becomes

$$I\ddot{\psi} = M - ml \begin{vmatrix} \sin \theta, & 0, & \cos \theta \\ \cos \theta \cos \psi, & \sin \psi, & -\sin \theta \cos \psi \\ a_x, & a_y, & a_z + g \end{vmatrix} \\ = M - ml[(a_z + g) \sin \theta \sin \psi + a_z \cos \psi - a_x \cos \theta \sin \psi]. \quad (B)$$

Equation (B) is of a general character. To take an important particular case, we assume that $\psi=0$ is a position of equilibrium and we consider the motion for small values of ψ . Further* we assume that $M = -\lambda\dot{\psi} - \mu\psi$, where λ and μ are constants, and we put

$$\frac{\lambda}{I} = 2A, \quad \frac{ml}{I} = B, \quad \frac{\mu}{I} = C, \quad n^2 = Bg \sin \theta + C.$$

Then the equation (B) becomes

$$\ddot{\psi} + 2A\dot{\psi} + [B\{(a_z + g) \sin \theta - a_x \cos \theta\} + C]\psi = Ba_y, \\ i. e. \quad \ddot{\psi} + 2A\dot{\psi} + [n^2 + B(a_z \sin \theta - a_x \cos \theta)]\psi = Ba_y. \quad (C)$$

A discussion of (C) is not possible without a knowledge of a_x , a_y , a_z as functions of t . For the purpose of the present investigation we assume that the motion of O relative to axes fixed in the earth is simple harmonic and we assume that

$$a_x = A_x \sin pt, \quad a_y = A_y \sin pt, \quad a_z = A_z \sin pt,$$

where A_x , A_y , A_z and p are constants, and we put $H = B(A_z \sin \theta - A_x \cos \theta)$.

Then (C) reduces to

$$\ddot{\psi} + 2A\dot{\psi} + (n^2 + H \sin pt)\psi = BA_y \sin pt. \quad (D)$$

Equation (D) belongs to the class of ordinary differential equations with periodic coefficients. H is a parameter and in many cases it can be assumed to be small. We shall first obtain the particular integral of (D) and afterwards its complementary function.

* The form assumed for M provides for damping and elastic forces.

§ 2. *The Particular Integral of (D).*

In order to reduce algebraic work we shall employ the artifice of using complex functions of a real variable. We therefore take the complex equation

$$\ddot{\psi} + 2A\dot{\psi} + (n^2 + H \sin pt)\psi = BA_y e^{ipt}, \quad \dots \quad (D')$$

and we take the imaginary part of the particular integral of (D') as the particular integral of (D). We put $\psi = \xi e^{ipt}$.

Then (D') becomes

$$\ddot{\xi} + (2ip + 2A)\dot{\xi} + (n^2 - p^2 + 2iAp + H \sin pt)\xi = BA_y. \quad \dots \quad (D'')$$

Now assume that $\xi = \xi_0 + H\xi_1 + H^2\xi_2 + \dots$,

where ξ_0, ξ_1, \dots are functions of t , and that the above series is convergent and can be differentiated term by term with respect to t as far as the second order time derivatives. When this substitution is made, (D'') becomes*

$$\begin{aligned} (\Sigma H^s \ddot{\xi}_s) + (2ip + 2A)(\Sigma H^s \dot{\xi}_s) + (n^2 - p^2 + 2iAp + H \sin pt)(\Sigma H^s \xi_s) \\ - BA_y = 0. \quad \dots \quad (1) \end{aligned}$$

Since (1) is valid for an arbitrary value of H in the range concerned, we equate to zero the coefficients of the different powers of H . By doing so we obtain the following sequence of equations, whose particular integrals have to be found:—

$$\ddot{\xi}_0 + (2ip + 2A)\dot{\xi}_0 + (n^2 - p^2 + 2iAp)\xi_0 = BA_y, \quad \dots \quad (2.0)$$

$$\ddot{\xi}_1 + (2ip + 2A)\dot{\xi}_1 + (n^2 - p^2 + 2iAp)\xi_1 + \sin pt \xi_0 = 0, \quad \dots \quad (2.1)$$

$$\ddot{\xi}_2 + (2ip + 2A)\dot{\xi}_2 + (n^2 - p^2 + 2iAp)\xi_2 + \sin pt \xi_1 = 0, \quad \dots \quad (2.2)$$

\vdots

$$\ddot{\xi}_s + (2ip + 2A)\dot{\xi}_s + (n^2 - p^2 + 2iAp)\xi_s + \sin pt \xi_{s-1} = 0. \quad \dots \quad (2.s)$$

\vdots

The particular integral of (2.0) is

$$\xi_0 = \frac{BA_y}{n^2 - p^2 + 2iAp}. \quad \dots \quad (3.0)$$

The equation (2.1) then becomes

$$\ddot{\xi}_1 + (2ip + 2A)\dot{\xi}_1 + (n^2 - p^2 + 2iAp)\xi_1 = -\frac{BA_y}{n^2 - p^2 + 2iAp} \sin pt,$$

whose particular integral is

$$\xi_1 = -\frac{1}{2i} \frac{BA_y}{n^2 - p^2 + 2iAp} \left[\frac{e^{ipt}}{n^2 - 4p^2 + 4iAp} - \frac{e^{-ipt}}{n^2} \right]. \quad \dots \quad (3.1)$$

* The summation symbol Σ is employed in (1) for the sake of brevity to cover the range of values 0, 1, 2, ... of s .

We can proceed in a similar manner for ξ_2, ξ_3, \dots . We stop, however, at the first power of H , and evaluate $(\xi_0 + H\xi_1)e^{ipt}$:

$$\begin{aligned} & (\xi_0 + H\xi_1)e^{ipt} \\ &= \frac{BA_y e^{ipt}}{n^2 - p^2 + 2iAp} - \frac{1}{2i} \frac{HBA_y}{n^2 - p^2 + 2iAp} \left[\frac{e^{2ipt}}{n^2 - 4p^2 + 4iAp} - \frac{1}{n^2} \right] \\ &= \frac{BA_y e^{i(pt - \epsilon_1)}}{\{(n^2 - p^2)^2 + 4A^2 p^2\}^{\frac{1}{2}}} - \frac{1}{2i} \frac{HBA_y e^{-i\epsilon_1}}{\{(n^2 - p^2)^2 + 4A^2 p^2\}^{\frac{1}{2}}} \left[\frac{e^{2ipt - i\epsilon_2}}{\{(n^2 - 4p^2)^2 + 16A^2 p^2\}^{\frac{1}{2}}} - \frac{1}{n^2} \right], \end{aligned}$$

where $\tan \epsilon_1 = \frac{2Ap}{n^2 - p^2}, \quad \tan \epsilon_2 = \frac{4Ap}{n^2 - 4p^2}.$

Hence, as far as the first power of H , the particular integral of (D) is

$$\begin{aligned} \psi &= \frac{BA_y \sin(pt - \epsilon_1)}{\{(n^2 - p^2)^2 + 4A^2 p^2\}^{\frac{1}{2}}} \\ &+ \frac{HBA_y}{2\{(n^2 - p^2)^2 + 4A^2 p^2\}^{\frac{1}{2}}} \left[\frac{\cos(2pt - \epsilon_1 - \epsilon_2)}{\{(n^2 - 4p^2)^2 + 16A^2 p^2\}^{\frac{1}{2}}} - \frac{\cos \epsilon_1}{n^2} \right]. \quad (P) \end{aligned}$$

§ 3. The Complementary Function of (D).

It will be useful first to transform the equation by a change of the variables and also the constants. Let

$$h = \frac{H}{p^2}, \quad N = \frac{\sqrt{n^2 - A^2}}{p}, \quad pt = \tau + \frac{\pi}{2} \quad \text{and} \quad \psi = e^{-At} \chi,$$

and let the prime (') be used to denote differentiation with respect to τ .

Then $\ddot{\psi} + 2A\dot{\psi} + (n^2 + H \sin pt)\psi = p^2 e^{-At} \{\chi'' + (N^2 + h \cos \tau)\chi\},$

and the equation for the complementary function becomes

$$\chi'' + (N^2 + h \cos \tau)\chi = 0. \quad (4)$$

For solving (4) we employ the method of Whittaker* and Ince†, which has been employed in some investigations.

h is the parameter which governs the solution. We assume that

$$\chi = e^{c\tau} u,$$

where

$$u = \cos(\nu\tau - \sigma) + \Sigma h^r P_r,$$

$$c = \Sigma h^r Q_r,$$

$$N^2 = \nu^2 + \Sigma h^r R_r,$$

where the summation is for positive integral values of r , the P_r 's being periodic functions of τ and Q_r 's and R_r 's being constants, and σ being a constant depending on the equation (4). No attempt is made below to

* E. T. Whittaker, "On the general solution of Mathieu's equation," Proc. Edin. Math. Soc. vol. xxxii. pp. 75-80 (1914).

† E. L. Ince, "On a general solution of Hill's equation," Monthly Notices, R. A. S. vol. lxxv. pp. 436-448 (1915).

evaluate σ . The conditions are that u is to be purely periodic and that the solution is to contain no term in $\sin(\nu\tau - \sigma)$.

$$u' = -\nu \sin(\nu\tau - \sigma) + \Sigma h' P_r',$$

$$u'' = -\nu^2 \cos(\nu\tau - \sigma) + \Sigma h'' P_r'',$$

and equation (4) becomes

$$u'' + 2cu' + c^2u + (N^2 + h \cos \tau)u = 0,$$

i. e.
$$-\nu^2 \cos(\nu\tau - \sigma) + \Sigma h'' P_r'' + 2c\{-\nu \sin(\nu\tau - \sigma) + \Sigma h' P_r'\} + c^2\{\cos(\nu\tau - \sigma) + \Sigma h' P_r'\} + (\nu^2 + \Sigma h'' R_r)\{\cos(\nu\tau - \sigma) + \Sigma h' P_r'\} + h \cos \tau \{\cos(\nu\tau - \sigma) + \Sigma h' P_r'\} = 0. \quad (5)$$

There is no term independent of h in the left-hand side of (5). Equating to zero the coefficient of h in the left-hand side of (5) we obtain

$$P_1'' - 2\nu Q_1 \sin(\nu\tau - \sigma) + \nu^2 P_1 + R_1 \cos(\nu\tau - \sigma) + \cos \tau \cos(\nu\tau - \sigma) = 0,$$

the solution of which is

$$Q_1 = R_1 = 0 \quad \text{and}$$

$$P_1 = \frac{\cos\{(\nu+1)\tau - \sigma\}}{2\{(\nu+1)^2 - \nu^2\}} + \frac{\cos\{(\nu-1)\tau - \sigma\}}{2\{(\nu-1)^2 - \nu^2\}},$$

provided $2\nu \neq 1$.

To the first order of h we may write $c=0$ and $N=\nu$, and the solution becomes

$$u = \cos(N\tau - \sigma) + \frac{1}{2} \left[\frac{\cos\{(N+1)\tau - \sigma\}}{(N+1)^2 - N^2} + \frac{\cos\{(N-1)\tau - \sigma\}}{(N-1)^2 - N^2} \right] h,$$

provided $2N \neq 1$.

Hence the complementary function of (D) is of the type

$$\psi = e^{-At} \left[\cos(N\tau - \sigma) + \frac{1}{2} \left\{ \frac{\cos\{(N+1)\tau - \sigma\}}{(N+1)^2 - N^2} + \frac{\cos\{(N-1)\tau - \sigma\}}{(N-1)^2 - N^2} \right\} h \right], \quad (6)$$

as far as the first power of h , provided $2N \neq 1$.

§ 4. Discussion of the Results.

In the solution of (D) in § 2 and § 3 above, h^2 and higher powers of h have been neglected. If terms in h^2 are taken into account we would obtain a closer approximation to the relation between the motion of O and the motion of the seismometer relative to its frame.

(1) First we examine the result in (P); for this purpose we consider the different terms in the right-hand side of (P).

(i) The term
$$\frac{BA_y \sin(pt - \epsilon_1)}{\{(n^2 - p^2)^2 + 4A^2 p^2\}^{\frac{1}{2}}}$$

is the one which alone is present when the correction due to the motion of O is not applied.

(ii) The term
$$\frac{HBA_y \cos(2pt - \epsilon_1 - \epsilon_2)}{\{(n^2 - p^2)^2 + 4A^2 p^2\}^{\frac{1}{2}} \{(n^2 - 4p^2)^2 + 16A^2 p^2\}^{\frac{1}{2}}}$$

represents an additional vibration superposed on the one due to the first term.

(iii) The term
$$-\frac{HBA_y \cos \epsilon_1}{2\{(n^2-p^2)^2+4A^2p^2\}^{\frac{1}{2}}n^2}$$

implies that, during the motion of O, the seismometer would have a constant deflection. Consequently, in a seismogram, the line of the mean positions in the oscillations would be parallel to, but different from, the trace, had the ground not been in motion. It must be mentioned, however, that in actual practice, the zero of a seismometer shifts sometimes due to instrumental defects and other causes and that it would be difficult to separate the shift of the zero due to the motion of the instrument relative to axes fixed in the earth from the possible shift, if any, due to other causes. This shift of the zero in some cases may make it possible to estimate the order of the ground acceleration, especially in the first movement, where the motion of O may be assumed to be an "unmixed oscillation."

We may perhaps be in a position to find some sort of criterion for the intensity of a shock from the shift of the zero, assuming, of course, that the spurious shift is absent or negligible.

From the displacement of the zero line, we can compute H, when n , A , p , etc., are known. From the actual reports about shocks, H can be correlated with the epicentral distance of the shock from the seismometer and the intensity. The intensity thus obtained would be free from the uncertainties due to the personal equation. This method would be specially valuable when $\theta=0$, as in the case of the Jaggar or the Wood-Anderson torsion instruments.

If $A=0$, *i. e.*, in the absence of damping, there would be resonance when $n=2p$ in addition to when $n=p$ (as in the usual theory). Even if A is different from zero but small, there would be a spurious magnification when $n=2p$ as well as when $n=p$. As regards the actual evaluation of the particular integral in case of the two resonance values of p , there is no difficulty; we have simply the "logarithmic" case.

(2) Next we examine the result in (6). There will be a "logarithmic" case when $2N=1$. If $A=0$, there will be resonance when $2n=p$; even if there be damping, there will be a resonance effect when $2N=1$, *i. e.*, if $n=\frac{1}{2}p$, if A is small.

For a teleseismic record, the ground motion is generally small and the above-mentioned effects may not be particularly noticeable. But when the ground motion is not small, as is the case when the seismometer is in the meizoseismal zone of an earthquake or when the earthquake is severe and its epicentre is comparatively near to the seismometer.

If h^2 and higher powers of h are considered, as it may be necessary to do when the ground motion for the seismometer is very severe, it is probable that there may be further resonance relations. All these make the choice of the free period of the seismometer a complicated task, especially in the case of strong motion.

§ 5. The "Berlage onset."

The assumption that the ground motion is simple harmonic is a first approximation to the actual motion in an earthquake pulse. Berlage* has suggested for the ground displacement the form $Ate^{-\alpha t} \sin pt$.

If we are to use in (C) the values for a_x, a_y, a_z derived from Berlage's form, the resulting equation would become much more complicated than (D). It is not considered in the present paper.

We may note, however, that if the terms containing a_x, a_y, a_z in the left-hand side of (C) are neglected, there will be factors

$$e^{-\Delta t} \quad \text{and} \quad e^{-\alpha t}$$

in the terms in the complementary function and the particular integral respectively. The question for consideration would then be how far it would be possible to ignore the complementary function in comparison with the particular integral, as is generally done in the case of a simple harmonic "onset" †.

LXXXIV. *Notices respecting New Books:*

Tables of Lagrangian Interpolation Coefficients prepared by the Mathematical Tables Project. [Pp. xxxvi+392.] (New York: Columbia University Press, 1944.) Price \$5.00.

SINCE a polynomial of degree n has $(n+1)$ coefficients, it may be adjusted to pass through $(n+1)$ assigned points. Lagrange showed long ago that a polynomial $y=f(x)$ to pass through the points (x_r, y_r) could be written down as a sum of terms of which the r th is

$$\frac{(x-x_1)(x-x_2)\dots(x-x_n)y_r}{(x_r-x_1)(x_r-x_2)\dots(x_r-x_n)},$$

in which the factor $(x-x_r)$ is omitted from the numerator and (x_r-x_r) naturally does not appear in the denominator. Thus the r th term reduces to y_r when $x=x_r$ and vanishes when x has any other of the values x_1, x_2, \dots, x_n . Consequently the sum of such terms, which is obviously of degree $(n-1)$, reduces to y_r for $x=x_r$.

In the practical use of this method for the approximate representation of a given function, the values of x_r will normally be equidistant with interval h , and the formula (with a "remainder" term of degree n) becomes

$$f(x) = \sum_k A_k^n(p) y_k + R_n(x) \quad \text{where } p \text{ is } \frac{x-x_0}{h}.$$

* Cf. Macelwane and Schon, "Introduction to Theoretical Seismology," Part 2, p. 103; also "Handbuch der Geophysik," Bd. IV, p. 369.

† It may be mentioned that the assumption that the pulse is not a mixture of more elementary oscillations is made for the sake of simplifying the work. If the pulse were assumed to be composed of a number of more elementary oscillations, the equations and their solutions would be correspondingly more complicated.

One of the commonest applications of the method is to approximate to the integral of the function $f(x)$, the integral from x_0 to x_0+ph being

$$h \sum M_k^n(p)y_k.$$

In this volume, values of A_k^n are given to at least 10 decimals for $n=3$ to 11 with p at intervals of 0.0001, 0.001, 0.01 or 0.1 according to its total range. The polynomials $M_k^n(p)$ are tabulated for $n=3$ to 11 at all values of k , and there are various subsidiary tables, together with a very valuable introduction and bibliography. J. H. A.

Table of Circular and Hyperbolic Tangents and Cotangents for Radian Arguments, by the Mathematical Tables Project. [Pp. xxxviii+410.] (New York : Columbia University Press, 1943.) Price \$5.00.

FROM the analyst's point of view, the inverse tangent, like the logarithm, has to be introduced in order to effect the integration of a rational function, in this case $1/(1+x^2)$. If the logarithm has been introduced, there is no need for a new function to integrate $1/(1-x^2)$, but there is a great gain in symmetry if one, the inverse hyperbolic tangent, is used. From this point of view it is really the tangent and hyperbolic tangent which are the inverse functions, and it is interesting to note that in fact the volume of arctan preceded the present volume in this series.

The function $\tanh x$, varying only from 0 to 1, presents no difficulties of interpolation, and thus $\coth x$ can be handled by interpolating $\tanh x$. In the case of $\tan x$ and $\cot x$, the same device can be used in the neighbourhoods of their respective infinities, but elsewhere the Gregory-Newton or Everett's formulæ are recommended; second central differences are given where necessary. Another method of interpolation near $x=0$ is to interpolate, not for $\cot x$ or $\coth x$ direct, but for $\cot x - \frac{1}{x}$ or $\coth x - \frac{1}{x}$. For inverse interpolation, as noted above, there is a volume in the same series, giving $\arctan x$, whilst, since $\arg \coth y$ is $\frac{1}{2} \log (y+1)/(y-1)$, the problem here is best solved with a table of natural logarithms.

The volume under review opens, after the explanatory matter, with a most interesting annotated bibliography, and then proceeds to the main table in which the four functions \tan and \tanh , \cot and \coth are tabulated side by side at intervals of 0.0001 from 0 to 2, to 8 figures. Then follows a short table from $x=0$ to 10 at intervals of 0.1, to 10 decimals, and finally a table for converting circular measure to radians or vice-versa, with tables of the coefficients in the interpolation formulæ mentioned earlier. The methods of calculation are such as to give high confidence in the results, and since the book has been produced by the photo-offset process, there will be no errors due to setting up the tables in type. J. H. A.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

LXXXV. *A Push-Pull Resistance-Capacity Coupled Oscillator.*

By W. F. LOVERING, B.Sc.*

[Received July 24, 1944.]

A RESISTANCE-CAPACITY coupled amplifier with a phase shifting network connected from output to input to provide positive feedback at one frequency will oscillate at that frequency, providing that the feedback voltage is sufficient. Several such oscillators have been described^{(1), (2)} but all are fundamentally similar; one or more valves may be used in the amplifier, the phase shifting network being then arranged to give the appropriate phase shift (0° , 180° , ... etc.) at the desired frequency. The frequency at which oscillation occurs is that at which the phase of the feedback is correct, and this may be controlled by variable resistors or condensers in the feedback circuit.

The phase of the feedback is mainly dependent on the constants of the feedback network and the frequency stability is, therefore, good, if stable resistors and condensers are used. However, the valve parameters have a small influence on the phase of the feedback and variations in these can affect the frequency. To minimise this effect it is usual to employ negative feedback; unbypassed cathode resistors may be used for this purpose.

The amplitude of oscillation adjusts itself to a value such that the effective amplification is just sufficient to maintain the oscillation. If the feedback voltage is greater than is necessary to maintain oscillation, the amplitude of oscillation increases, and the valves operate over a greater portion of their characteristics so that characteristic curvature reduces the average Gm. during the cycle. This process continues until the effective amplification just counterbalances the attenuation in the feedback circuit. If the amplification is only just sufficient the wave-form of the oscillation will be practically sinusoidal; an excessive amplification results in operation over an appreciably curved characteristic so that the output is distorted, containing mainly even harmonics. If the amplitude of oscillation is allowed to become so large that the valve runs into grid current, and also operates over the curved part of its characteristic, the output wave-form is flattened on both peaks, *i. e.*, it contains mainly odd harmonics. In an extreme case the flow of grid current can convert the oscillation into a relaxation oscillation.

The application of negative feedback increases the signal which the amplifier can handle without appreciable distortion, and makes the

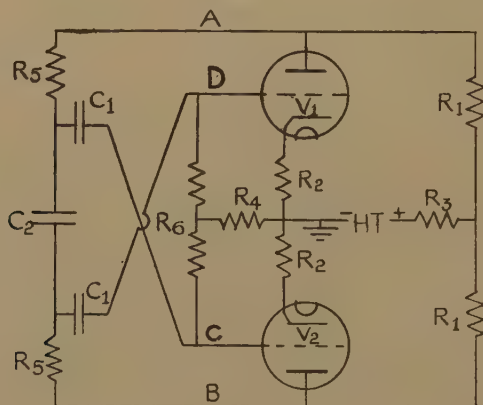
* Communicated by Professor L. F. Bates, Ph.D., D.Sc.

amplification and therefore the wave-form less dependent on the valve characteristics.

One common source of trouble with this type of oscillator is feedback via a common impedance in the H.T. circuit. This may be negative or positive or may introduce phase shifts, depending on the oscillator and H.T. used. Where an amplifier is operated from the same H.T. supply this effect may cause the frequency and the amplitude of the oscillation to vary with the setting of the output control. Elaborate decoupling is usually necessary to minimise the bad effects of the H.T. source impedance.

Three-phase oscillators have been described using three valves, each handling a signal 120° out of phase with that handled by the other valves^{(3), (4)}. With these circuits there is no alternating component in the current drawn from the H.T. source and the effects of supply impedance are thus removed. Though useful for some purposes, these oscillators are rather complicated and hard to adjust.

Fig. 1.



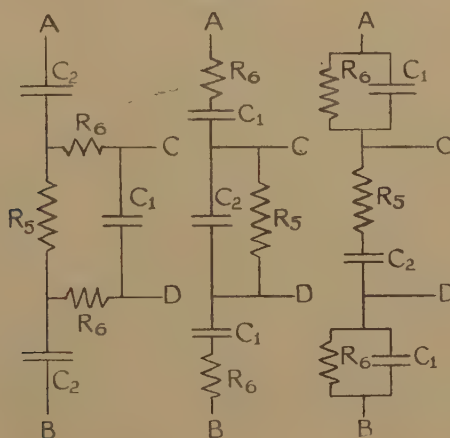
This article describes a simple two-valve oscillator which, by the use of the push-pull connection, provides freedom from H.T. impedance troubles and at the same time gives the improvement in wave-form which is usual with push-pull circuits.

The circuit of the oscillator is shown in fig. 1. The feedback circuit shown is that thought by the author to be the best of those he has considered. Alternative feedback circuits are shown in fig. 2. These various circuits are all such that the voltage between C and D is in phase with that between A and B at the frequency at which $\omega^2 C_1 C_2 R_5 R_6 = 1$; at other frequencies the voltage between C and D leads, or lags, that between A and B. The grid of V_1 is connected to point D, and the grid of V_2 to point C, so that the P.D. applied between the two grids is in antiphase with that between the two anodes at the frequency given by the above expression.

The pairs of resistors R_1 , R_2 and R_5 and the pair of condensers C_1 should preferably be matched. If the matching is exact the centre point of R_6

(made up of two equal resistors) will be at earth potential and may be joined directly to earth. Similarly, the junctions of the two anode loads may be joined directly to H.T. positive. However, it is difficult in practice to make the matching exact and the resistances R_3 and R_4 will usually improve the performance. For example, if V_1 carries a larger alternating anode current than V_2 the difference current will flow in R_3 . This will raise the alternating P.D. between A and earth and lower that between B and earth, so that the alternating P.D. between C and earth (*i. e.* the input signal to V_2) will rise and the input to V_1 will fall. This action tends to maintain equality between the signals applied to the two valves and so ensures a good wave-form; R_4 performs a similar function. In addition R_3 may be made large compared with the internal impedance of the H.T.

Fig. 2.



source, when variation in the latter will have a negligible effect on the oscillator.

The feedback circuit.—It is convenient to consider the amplifier and feedback circuit separately. For convenience in calculation the symmetrical circuits shown in figs. 1 and 2 may be replaced by the circuits of fig. 3. For example, circuit 3 (a) replaces the circuit shown in fig. 1; in this case $R_B = 2R_5$, $R_A = R_6$, $C_B = C_2$ and $C_A = \frac{C_1}{2}$; similar transformations are made in the other cases.

The feedback network is required to have such characteristics that the output voltage E_0 is in phase with the input voltage E_i at one frequency only. The oscillator oscillates at this critical frequency (f_0), and any tendency for the frequency of oscillation to change is offset by the resulting phase shift between E_i and E_0 . The greater the change in the phase angle (ϕ) between E_0 and E_i for a given change of frequency, therefore, the better the frequency stability. The smaller the attenuation in the filter

at the critical frequency, the better, since less amplification is required and therefore more negative feedback may be employed.

The criteria of a good feedback network, therefore, are a large value of $\frac{d\phi}{d\omega}$ (or $\frac{d \tan \phi}{d\omega}$), and a large value of $\frac{E_0}{E_i}$, at the critical frequency.

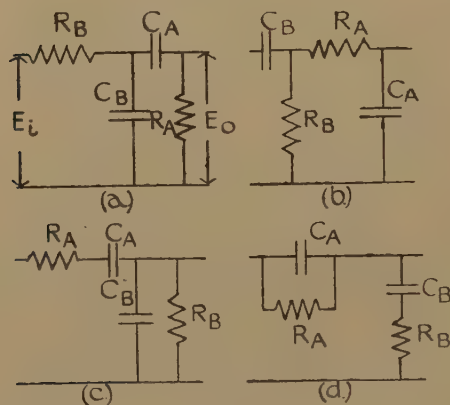
Consider the circuit of fig. 3 (a).

$$\frac{E_0}{E_i} = \frac{R_A}{R_A - j/\omega C_A} \cdot \frac{-j/\omega C_B (R_A - j/\omega C_A)}{R_A - j/\omega C_A - j/\omega C_B} \cdot \frac{1}{R_B - \frac{j/\omega C_B (R_A - j/\omega C_A)}{R_A - j/\omega C_A - j/\omega C_B}} \quad (i)$$

which simplifies to

$$\frac{E_0}{E_i} = \frac{1}{1 + R_B/R_A + R_B^2 C_B/R_A C_A + j(\omega C_B R_B - 1/\omega C_A R_A)} \quad (ii)$$

Fig. 3.



The phase angle between E_0 and E_i is given by

$$\tan \phi = \frac{1 - \omega^2 C_A C_B^2 R_A R_B}{\omega (R_A C_A + R_B C_B + R_B C_A)} \quad (iii)$$

E_0 and E_i are in phase when

$$\omega_0^2 = (2\pi f_0)^2 = \frac{1}{C_A C_B R_A R_B} \quad (iv)$$

Substituting $R_A C_A = x^2$, $R_B C_B = 1/x^2$, $C_B = n C_A$, and $\omega_0 = \sqrt{C_A C_B R_A R_B}$ in equations (ii) and (iii) gives

$$\frac{E_0}{E_i} = \frac{n x^2}{n x^2 + n + 1 + j n x A} \quad \text{where } A = \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \quad (v)$$

and $\tan \phi = \frac{-n x A}{n x^2 + n + 1}; \quad (vi)$

by differentiating (vi) and substituting $\omega = \omega_0$ in the resulting expression

$$\frac{d \tan \phi}{d\omega} = \frac{-2 n x}{\omega_0 (n x^2 + n + 1)} \quad \text{at the oscillation frequency.} \quad (vii)$$

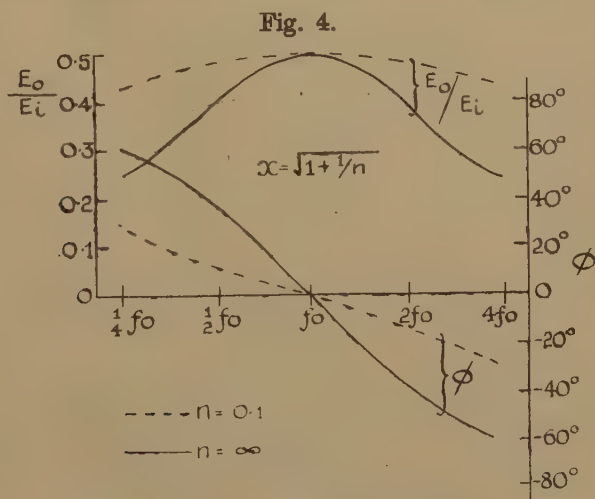
Examination of this expression shows that $\frac{d \tan \phi}{d\omega}$ has its maximum possible value of $\frac{-1}{\omega_0}$ when $n \rightarrow \infty$ and $x=1$, i. e., when $C_A R_A$ is equal to $C_B R_B$ and R_A is much greater than R_B .

For any given value of n , $\frac{d \tan \phi}{d\omega}$ is maximum when $x = \sqrt{1+1/n}$. Substituting this value for x in equations (v) and (vii) shows that

$$\frac{E_0}{E_i} = 0.5 \quad \text{and} \quad \frac{d \tan \phi}{d\omega} = \frac{-1}{\omega_0 \sqrt{1+1/n}}$$

at the critical frequency ($A=0$).

It is clearly desirable to make n as large as possible and in an oscillator meant for use at one frequency only this presents no difficulty. However, if the oscillation frequency is to be varied by varying either resistance or



capacity values it will be necessary to gang together either C_A and C_B or R_A and R_B . If one component only is varied then the ratio $\frac{E_0}{E_i}$ at the critical frequency will also vary (equations ii and v). This will result either in an increase of amplitude of oscillation, and consequent poor wave-form if $\frac{E_0}{E_i}$ increases, or in a cessation of oscillation if $\frac{E_0}{E_i}$ decreases. This can be overcome of course by readjustment of the valve anode load or cathode resistance to make the amplification suit the changed value of $\frac{E_0}{E_i}$, but this method is laborious. Ganged components make it possible to change the product $C_A C_B R_A R_B$ without changing the ratio $\frac{E_0}{E_i}$ at the critical frequency. Ganged condensers or resistors can be used, convenient values of n being $\frac{1}{2}$, 1 or 2.

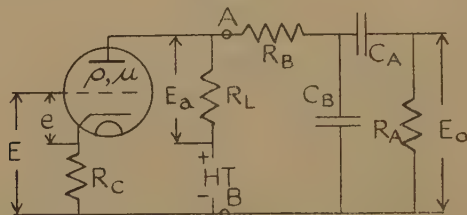
With $n=1$, $\frac{d \tan \phi}{d\omega}$ has a maximum value of $-0.707/\omega_0$ when $x=\sqrt{2}$.

With $n=2$, $\frac{d \tan \phi}{d\omega}$ has a maximum value of $-0.82/\omega_0$ when $x=\sqrt{1.5}$.

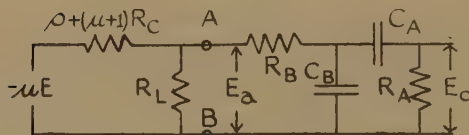
It will be seen that the choice of a value for n is not critical; similarly x may depart appreciably from its optimum value without any considerable effect on the performance of the network.

All the circuits of fig. 3 have been analysed in the above manner and the results of the analysis tabulated in Tables I. and II., from which it will be seen that the circuit of fig. 3 (a) is the best, giving the least attenuation for a given value of $\frac{d \tan \phi}{d\omega}$. Fig. 4 shows how $\frac{E_0}{E_i}$ and ϕ for this circuit change as ω is changed.

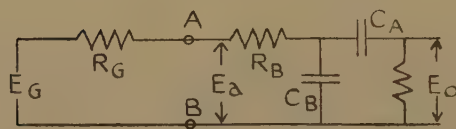
Fig. 5.



(a).



(b).



(c).

The Amplifier.—As with the feedback network it is more convenient to consider one half only of the push-pull circuit; the equivalent single-ended circuit is shown in fig. 5 (a). If the impedance of the valve anode load between A and B is Z_L , then the total impedance of the external anode to cathode circuit is $R_C + Z_L$. If e is the signal applied between grid and cathode of the valve then the P.D. developed between cathode and earth is $-\mu e R_C / (\rho + R_C + Z_L)$ and the P.D. E applied between grid and earth must be $\left(\frac{\mu R_C}{\rho + R_C + Z_L} + 1 \right) e$ volts, i. e.,

$$E = \left(\frac{\rho + (\mu + 1) R_C + Z_L}{\rho + R_C + Z_L} \right) e \text{ volts.} \quad \dots \dots \dots \text{(viii)}$$

TABLE I.

Characteristics of the filters of figs. 1, 2 and 3.

	Circuit 3 a.	Circuit 3 b.	Circuit 3 c.	Circuit 3 d.
$\frac{E_0}{E_i}$	$\frac{nx^2}{nx^2+n+1+jnxA}$	$\frac{n}{nx^2+1+n+jnxA}$	$\frac{1}{nx^2+1+n+jnxA}$	$\frac{x^2+1+jxA}{nx^2+x^2+1+jxA}$
$\tan \phi$	$\frac{-nxA}{nx^2+n+1}$	$\frac{-nxA}{nx^2+n+1}$	$\frac{-nxA}{nx^2+n+1}$	$\frac{nxA}{\left(x+\frac{1}{x}+nx\right)\left(x+\frac{1}{x}\right)+A^2}$
$\frac{d \tan \phi}{d\omega}$ at the critical frequency	$\frac{-2nx}{\omega_0(nx^2+n+1)}$	$\frac{-2nx}{\omega_0(nx^2+n+1)}$	$\frac{-2nx}{\omega_0(nx^2+n+1)}$	$\frac{2nx^3}{\omega (nx^2+x^2+1)(x^2+1)}$
Optimum value for x	$\sqrt{1+\frac{1}{n}}$	$\sqrt{1+\frac{1}{n}}$	$\sqrt{1+\frac{1}{n}}$	$\sqrt{\frac{n+2+\sqrt{n^2+16n+16}}{2(n+1)}}$
$\frac{R_A}{R_B}$ to give optimum x	$n+1$	$n+1$	$n+1$	—
Best value for n	∞	∞	∞	∞
$n = \frac{C_B}{C_A}; \quad x = \sqrt{\frac{C_A R_A}{C_B R_B}}; \quad A = \frac{f}{f_0} - \frac{f_0}{f}; \quad f_0 = \text{critical frequency.}$				

The P.D. (E_A) developed between anode and earth is given by

$$E_a = \frac{-\mu e Z_L}{\rho + R_C + Z_L}, \quad \dots \dots \dots (\text{ix})$$

whence

$$E_a = \frac{-\mu E Z_L}{\rho + (\mu + 1)R_C + Z_L} \quad \dots \dots \dots (\text{x})$$

TABLE II.

Comparison of performance of filters when x is given its optimum value.

		$\frac{E_0}{E_i}$	$\frac{d \tan \phi}{d\omega}$
Values with x having its optimum value	Circuit 3a	0.5	$-1/\omega_0 \sqrt{1+1/n}$
	„ 3b	$0.5/(1+1/n)$	„
	„ 3c	$0.5/1+n$	„
	„ 3d	—	—
Optimum x $n=0.5$	Circuit 3a	0.5	$-0.576/\omega_0$
	„ 3b	0.17	„
	„ 3c	0.33	„
	„ 3d	0.74	$+0.238/\omega_0$
Optimum x $n=1$	Circuit 3a	0.5	$-0.707/\omega_0$
	„ 3b	0.25	„
	„ 3c	0.25	„
	„ 3d	0.59	$+0.378/\omega_0$
Optimum x $n=2$	Circuit 3a	0.5	$-0.82/\omega_0$
	„ 3b	0.33	„
	„ 3c	0.17	„
	„ 3d	0.435	$+0.53/\omega_0$
Optimum x $n=\infty$	Circuit 3a	0.5	$-1/\omega_0$
	„ 3b	0.5	„
	„ 3c	0.0	„
	„ 3d	0.0	$+1/\omega_0$

Thus the valve, with its cathode resistance, acts as a generator having an E.M.F. of $-\mu E$ volts, and an internal resistance of $(\rho + (\mu + 1)R_C)$ ohms. The circuit with the valve replaced by the equivalent generator is shown in fig. 5 (b).

Consider the case when the input impedance of the filter is very much greater than R_L , so that Z_L is practically equal to R_L , then

$$\text{stage gain} \quad \frac{E_a}{E} = \frac{\mu R_L}{\rho + (\mu + 1)R_C + R_L} \quad \dots \dots \dots (\text{xi})$$

and E_a is in antiphase with E . This equation may be written as

$$\frac{E_a}{E} = \frac{R_L}{R_C} \div \left(1 + \frac{\rho + R_L + R_C}{\mu R_C} \right), \quad \dots \dots \dots (\text{x i})$$

or the stage gain is approximately equal to $\frac{R_L}{R_C}$ if μR_C is large compared

with $\rho + R_L + R_C$. That is, if a valve with a high amplification factor is used and R_C and R_L are made large the stage gain becomes almost independent of the valve characteristics, and also, therefore, of the supply voltages. It may be noted that push-pull connection makes the anode load, and thus the stage gain, independent of the impedance of the H.T. supply.

The output voltage from the filter is the input voltage to the other valve, so that for oscillation E_0 should be equal to E ; that is, if the filter of fig. 3 (a) is used and the optimum conditions are employed, the valve should provide a stage gain of two times. Whichever filter is used the stage gain should be such as just to offset the attenuation in the filter.

If the input impedance of the filter is not very much greater than R_L , then the effective anode load Z_L will be a capacitive impedance; the ratio $\frac{E_a}{E}$ will be reduced and E_a will not be in antiphase with E . The result

of this will be that oscillation will not occur at the frequency at which E_0 is in phase with E_a (i. e., when $\omega^2 R_A R_B C_A C_B = 1$), but at a lower frequency such that E_0 is in antiphase with E and leads E_a . For this case a more exact analysis of the circuit of fig. 5 (b) is necessary.

The circuit is most easily analysed with the aid of Thévenin's Theorem; the portions of the circuit to the left of the two points A and B in fig. 5 (b) are replaced by an equivalent generator, as shown in fig. 5 (c).

The E.M.F. (E_G) of the equivalent generator is given by

$$E_G = \frac{-\mu E R_L}{\rho + (\mu + 1) R_C + R_L} \quad \dots \dots \dots \text{(xiii)}$$

and its internal resistance (R_G) by

$$R_G = \frac{R_L [\rho + (\mu + 1) R_C]}{R_L + \rho + (\mu + 1) R_C} \quad \dots \dots \dots \text{(xiv)}$$

Inspection of this circuit shows that E_0 is in phase with E_G and therefore in antiphase with E when

$$\omega^2 (R_G + R_B) R_A C_A C_B = 1, \quad \dots \dots \dots \text{(xv)}$$

and that at this frequency $E_0 = E_G/2$ if $R_A C_A / (R_G + R_B) C_B$ has its optimum value, i. e., the circuit is the same as that of fig. 3 (a) but with $(R_G + R_B)$ taking the place of R_B .

For E_0 to be equal in magnitude to E , $\frac{\mu R_L}{\rho + (\mu + 1) R_C + R_L}$ must be equal to two; this is the same condition as specified above in the simplified analysis.

The solution of the complete circuit by the classical method is longer than that given above and is less illustrative; it is given for comparison in an appendix

The effect of the filter impedance in parallel with R_L is to reduce the frequency of oscillation and also, since R_G includes the valve parameters, the frequency is made to depend to some extent upon the valve characteristics. However, rewriting equation (xiv) and substituting

$$\mu R_L / [\rho + (\mu + 1)R_G + R_L] = 2$$

gives

$$R_G = R_L \div [1 + 2/(\mu - 2)], \quad \dots \dots \dots (xvi)$$

from which it may be seen that if μ is large then R_G is practically equal to R_L and is very little affected by any change in the valve characteristics. In addition, R_G is in series with R_B and may be made only a fraction of the total resistance $R_G + R_B$, so that changes in the valve will have very little influence on the frequency. In choosing circuit values R_L should not be made too large since this would make R_G a large proportion of $R_G + R_B$. Also, since R_L carries the steady anode current of the valve, its value is more likely to change due to heating than is that of R_B .

Effect of Anode-Grid Capacity.—Valve capacities have been ignored in the above analysis and if pentodes are used the effects of these capacities will usually be negligible. The anode-cathode and grid-cathode capacities of a triode are small and in general may be ignored, but the anode-grid capacity, because of its position in the circuit, may exert an appreciable influence. The analysis of the circuit including anode-grid capacity is laborious and will not be given here (the method of the solution is given in Appendix II.), but the results may be quoted.

The condition for oscillation is

$$A \left(= \frac{\mu R_L}{\rho + (\mu + 1)R_G + R_L} \right) \geq \frac{(R_A + R_B + R_G)C_A + (R_A + R_G)C_C + (R_G + R_B)C_B}{R_A(C_A - C_C)}$$

approximately, where C_C is the anode to grid capacity.

This may be written as

$$A \geq \left(\frac{nx^2 + n + 1}{nx^2} + \left(\frac{R_G}{R_A} + 1 \right) \frac{C_C}{C_A} \right) \div \left(1 - \frac{C_C}{C_A} \right), \quad \dots \dots (xvii)$$

where $n = C_B/C_A$ and $x^2 = R_A C_A / (R_G + R_B) C_A$.

If x is given its optimum value, $\frac{nx^2 + n + 1}{nx^2}$ is equal to two.

Thus the presence of the anode to grid capacity necessitates an increased amplification.

The oscillation occurs at a frequency given by

$$\frac{1}{\omega^2 C_A C_B C_C} = \frac{(R_B^2 + 4R_G R_A + 2R_B R_A)}{C_B} + \frac{(R_B^2 + R_A R_G + R_A R_B + 2R_G R_B)}{C_A} \\ + \frac{R_A (R_G + R_B)}{C_C} + (R_B^2 + R_G R_B) \frac{C_B}{C_A^2} + \frac{R_G R_B C_C}{C_A^2} \text{ approximately,} \\ \dots \dots \dots (xviii)$$

from which it may be seen that the presence of C_C reduces the frequency of oscillation. If C_A and C_B are large compared with C_C this expression is

very little different from equation (xv), and for most purposes equation (xv) is sufficiently accurate. The interest in this expression lies in the fact that it reveals the possibility of oscillation if either C_B and/or R_B is zero. If both R_B and C_B are zero the circuit (fig. 6) is practically that of the multivibrator. It is known that multivibrators using triodes cease to function normally when the coupling condensers are made small to obtain a high oscillation frequency, and this analysis indicates the reason for this; under such conditions the action resembles that of a resistance capacity coupled oscillator.

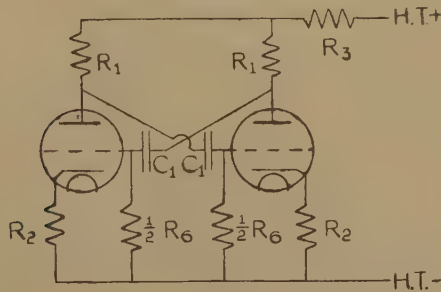
If both R_B and C_B are zero the condition for oscillation becomes

$$\frac{\mu R_L}{\rho + (\mu + 1)R_C + R_L} \geq \frac{R_A + R_G}{R_A} \cdot \frac{C_A + C_C}{C_A - C_C}, \dots \dots (xix)$$

and the frequency is given by

$$4\omega^2 C_A C_C R_A R_G = 1. \dots \dots \dots (xx)$$

Fig. 6.

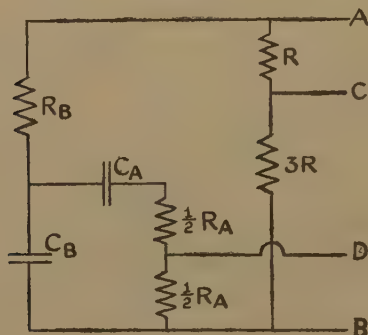


It is possible to employ circuits in which R_B and/or C_B is zero, but examination of equation (xviii) shows that it is better to make R_B and C_B comparatively large so that variations in R_G and in the anode to grid capacity have a negligible influence on the frequency of oscillation. Oscillators for the higher audio-frequencies should preferably use pentodes instead of triodes. The two screen grids should be connected to H.T. via a common resistance; the two suppressor grids should be earthed.

Output Connection.—Although this oscillator may be used to supply power directly to a load, such an arrangement is bad. As with all oscillators variations in the load will vary the frequency. A buffer stage should be interposed between the oscillator and the load. A single valve output may be used, but a symmetrical output wave-form containing no even harmonics is obtained if a push-pull output stage is used. Since the input impedance of a triode valve varies with its anode load, it is better to use pentode amplifiers or cathode follower stages following the oscillator. The output from the oscillator may be taken from one or both anodes, or from the cathodes; the former gives a higher output voltage; the latter gives greater freedom from any effects of load variation.

Alternative Feedback Circuits.—Although the symmetrical circuits of figs. 1 and 2 are preferable, they may be inconvenient if a wide frequency range is required. As has been shown, the ratio $C_1 : C_2$ or $(R_G + R_B) : R_A$ should remain constant as either C or R is changed. No standard ganged condensers which may be connected as shown are available, while variation of R requires four ganged resistors, of which two (R_B) must have an awkward law to allow for the presence of R_G . The difficulty may be overcome if one valve only is fed by a selective filter and the other receives its input from a plain potential divider arranged to give the same amplitude input signal. One such feedback circuit is illustrated in fig. 7. V_1 is fed via the filter, $R_B C_B C_A R_A$, while V_2 is fed by the potential divider R, $3R$. Each valve receives an input signal equal to one-quarter of the anode to anode P.D., providing that $R_A/R_B = C_B/C_A + 1$. The grid of V_2 must be isolated from the H.T. by a large condenser connected to C and must be earthed through a high resistance grid leak. If this circuit is used

Fig. 7.



R_B must be very much greater than R_G , otherwise, as reference to fig. 5 (c) and fig. 7 will show, the presence of R_G will affect the phase relations. With this or other similar circuits C_A and C_B may be sections of a standard ganged condenser. It is preferable to avoid the use of such circuits, since it is difficult in practice to preserve exact symmetry and the performance may be comparatively poor.

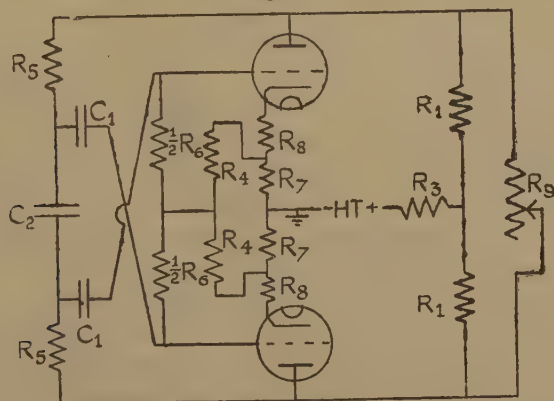
Experimental Results.—Several of the points raised above require experimental confirmation; possibly the most important was the explanation given of the reason for poor wave-form. One author⁽³⁾ has suggested that the phase shifting network controls the purity of wave-form, since harmonics suffer a large phase shift in the network and do not therefore receive positive feedback.

Other points to be investigated were the frequency stability and the influence of supply voltages in an oscillator with the normal circuit, the possibility of oscillation with C_B and/or R_B zero, and the highest and lowest frequency at which the oscillator would function.

Most of the experiments were carried out using the circuit of fig. 8, where $R_1=50,000$ ohms, $R_3=25,000$ ohms, $R_4=0.1$ megohms, $R_7=10,000$ ohms, $R_8=2,000$ ohms and $R_6=1$ megohm; the valves and the values of the other components were changed in different experiments. In this circuit the resistances R_8 were used to provide the correct value of steady grid bias, as the total cathode resistances had been chosen to give the required A.C. negative feedback and were greater than was desirable for the bias resistances. The two resistances R_4 replaced the single resistance R_4 in fig. 1. No standard condensers or resistors were available, so ordinary radio components were used in all the experiments. It was expected that some frequency drift might result from the fact that such components usually have appreciable temperature coefficients.

Wave-form.—The explanation given of the reason for poor wave-form was confirmed. Every oscillator made, no matter what feedback circuit

Fig. 8.



was employed, gave a very good wave-form as long as the amplification was only just sufficient to maintain oscillation. An increase in amplification, whether arising from an increase in anode load, a reduction of the cathode resistance or any other cause, resulted in a distorted wave-form, the distortion increasing as the amplification was raised. A reduction in the attenuation of the filter without a corresponding reduction in amplification caused the same effect.

Frequency Measurement.—The frequency was measured by comparison with 50 cps. time-controlled mains. The 50 cps. supply was applied to the X-plates of a C.R.O. and a small 50 cps. signal in quadrature with the X-plate supply was also applied to the Y-plates; this gave an elliptical trace on the C.R.O. screen. The output from one anode of the oscillator was applied, via an amplifier, to the Y-plates, and the oscillator frequency was adjusted to give a stationary trace. The elliptical trace separated the go and return traces of the spot and so simplified the counting of peaks in the pattern. Any variation of frequency caused the pattern to rotate,

the number of peaks passing a point on the screen in one second giving the slip frequency. The characteristics of four oscillators were measured in this way.

Oscillator No. 1.—Valves type M.H.4. $C_A=250\ \mu\mu\text{ F.}$, $C_B=1100\ \mu\mu\text{ F.}$, $R_5=100,000\ \text{ohms}$ approximately, $R_9=250,000\ \text{ohms}$ approximately. The resistances R_5 and R_9 were adjusted so that with an H.T. voltage of 210 volts the oscillator gave a good output wave-form with a frequency of 500 cps.

The frequency was found to change, at first rapidly and then more slowly, after switching on; a total change of 1 to 2 per cent. occurred in about two hours. It was considered that this change was mainly due to thermal changes in component valves. To test this, the oscillator was run for two hours and then the hot valves were replaced by cold valves. The frequency was then measured from two minutes after oscillation commenced. In the first fifteen minutes the frequency fell by one cycle per second, and by another 0.5 cps. in the next twenty minutes; thereafter it remained constant within 0.2 cps.

It appeared probable from the observations that much of the variation of ± 0.2 cps. was due to variations of the mains frequency, and not to that of the oscillator. The pattern on the screen would remain stationary for minutes, then suddenly would begin to drift at up to 0.2 cps., would again become stationary, and so on. The changes occurred at random and it seems more likely that they were due to the switching on and off of loads on the mains than to changes in the oscillator. This conclusion was supported by the fact that the operation of a nearby lift was accompanied by a slight drift of the pattern on the C.R.O. screen.

When the oscillator was thoroughly warmed up the H.T. and L.T. voltages were changed in turn and the resulting frequency change measured. A change of H.T. voltage from 210 v. to 100 v. resulted in a 1 cps. fall in frequency; while a change of L.T. voltage from 4 v. to 3 v. caused a fall of less than 0.2 cps.

Changing valves at random for others of the same type caused a frequency shift of less than 2 cps., while even substitution of M.H.L.4 valves ($\mu=20$) for M.H.4 valves ($\mu=40$) caused a change of only 3 per cent.

Oscillator No. 2.—Valves type M.H.L.4. $R_9=450,000\ \text{ohms}$ approximately, other values as in Oscillator No. 1. Frequency 500 cps.

This oscillator was practically as stable as No. 1, in spite of the lower amplification factors of the valves.

Drift after initial warming up period less than 0.2 cps. (0.04 per cent.).

H.T. changed from 210 v. to 120 v.—frequency fell by 1 cps. (0.2 per cent.).

L.T. changed from 4 v. to 3 v.—frequency fell by 1.5 cps. (0.3 per cent.).

Oscillator No. 3.—Valves type M.H.4. $R_5=73,000\ \text{ohms}$ approximately, $C_2=200\ \mu\mu\text{ F.}$, $C_1=0.01\ \mu\text{ F.}$, $R_6=300,000\ \text{ohms}$, $R_9=84,000\ \text{ohms}$ approximately, $f=500\ \text{cps.}$

This circuit was used to investigate the effect of using a feedback network with component values far from the optimum. With the valves used $n=1/25$, $x=6.1$. This gives $E_0/E_i=0.6$, so that an amplification of only 1.66 times was required; this accounts for the low value of R_9 . $\frac{d \tan \phi}{d\omega}$ for this circuit was only $-0.19/\omega_0$ as compared with the value of $-0.9/\omega_0$ obtained with the feedback network used in Oscillators Nos. 1 and 2.

This oscillator was rather more critical in adjustment than Nos. 1 and 2, and the output voltage was very dependent upon the H.T. voltage, oscillation ceasing when the H.T. voltage was reduced to 140 v.

Frequency shift after warming up less than 0.3 cps. (0.06 per cent.).

H.T. changed from 210 v. to 160 v.—frequency fell by 0.5 cps. (i.e. about the same proportional change as with the other oscillators).

Oscillator No. 4.—Valves type M.H.4. $R_1=\frac{1}{2}$ megohms, $R_3=25,000$ ohms, $R_7=100,000$ ohms, $R_4=100,000$ ohms, $R_8=6,000$ ohms, $R_6=1$ megohm, $C_1=200 \mu\mu$ F., $C_2=500 \mu\mu$ F., $R_5=38,000$ ohms approximately, $f=750$ cps.

With these entirely different circuit values the frequency stability was quite as good, the drift being less than 0.3 cps. A change of H.T. voltage from 210 v. to 100 v. caused a fall of frequency of 2 cps. (0.26 per cent.).

Voltage Output Characteristics.—The output voltage of all these oscillators was found to be very dependent upon the H.T. voltage, falling rapidly as the H.T. voltage was reduced. This was to be expected from the manner in which the amplitude of oscillation adjusts itself. Fig. 9 shows the output voltage characteristics of Oscillators 1 and 2. Terman has described a negative feedback circuit which may be used in resistance capacity coupled oscillators to minimise variations in output voltage⁽¹⁾.

The characteristics shown in fig. 9 refer to a particular set of operating conditions. An increase in the value of R_9 , for example, would enable oscillation to be maintained with much lower H.T. voltages, but would result in poor wave-forms with high H.T. voltages. The adjustment of amplification should be carried out to give good wave-form over the range of H.T. voltages it is desired to use.

Oscillator No. 5.—This was obtained by short-circuiting the resistances R_5 in Oscillator No. 1 for the purpose of testing conclusions drawn from equation (xviii).

From the values given the frequency of Oscillator No. 1 should have been approximately 582 cps. if the anode-grid capacities were ignored, or 570 cps. with an anode-grid capacity of $6 \mu\mu$ F. (Referring to fig. 5(c) and fig. 8, $R_G \div \frac{2R_1R_9}{2R_1+R_9}$, $R_B=2R_5$, $C_2=C_B$, $C_1=\frac{C_A}{2}$, $R_6=R_A$.) Since the condensers had tolerances of ± 15 per cent. from their nominal values, the actual frequency of 500 cps. is reasonably close to the calculated values. Making R_B zero should raise the frequency in the ratio

$$\sqrt{(R_G+R_B)/R_G} : \text{or } 1.96 : 1,$$

that is, to 980 cps. The measured frequency was just less than 1000 cps. The circuit of this oscillator is shown in fig. 10, which shows its resemblance to that of the Wynn Williams counting circuit.

Oscillator No. 6.—High frequency limit. In this oscillator C_2 was removed (fig. 10) and the condensers C_1 were made $100\ \mu\mu$ F. variable. Apart from the provision of steady bias to the two grids the circuit is the same as that of a multivibrator. As in the other oscillators the amplification could be controlled by changing the value of R_9 and thus changing the value of the anode loads ($2R_1$ in parallel with R_9).

Fig. 9.

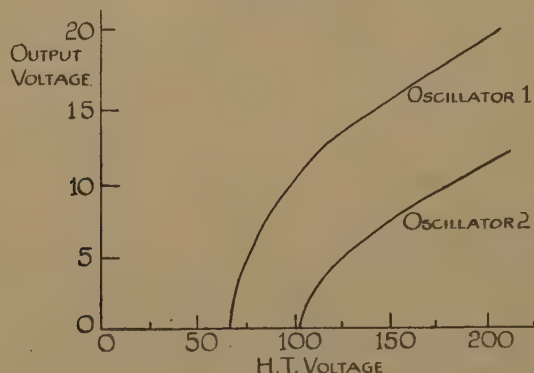
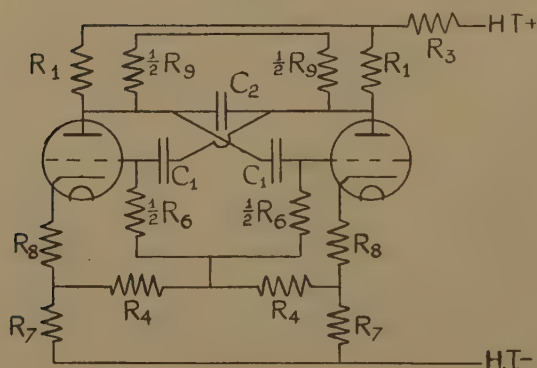


Fig. 10.



It was found that when the amplification was more than sufficient to maintain oscillation the normal multivibrator wave-form was produced. As the amplification was reduced the period of the relaxation oscillation became shorter and the corners of the wave were rounded off. Finally, with an amplification just sufficient for oscillation, the wave-form became sinusoidal.

Reduction of C_1 resulted in an increase in the frequency of the oscillation and necessitated an increase in amplification, thus verifying the conclusions of equations (xix) and (xx). The upper limit of oscillation with

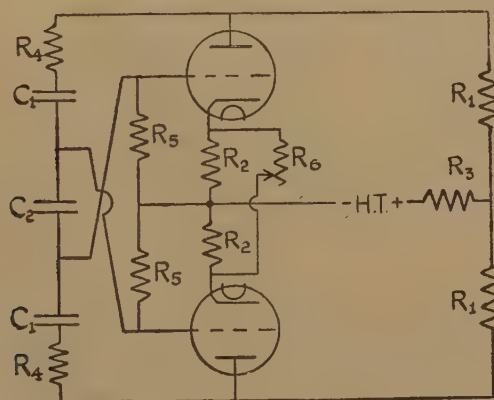
this oscillator was obtained with $C_1 = 20 \mu\text{F}$. and with R_9 open circuited; the frequency then was 26,000 cps. The frequency calculated from equation (xx) with these circuit values and $C_0 = 6 \mu\text{F}$. is 32,500 cps.—a sufficiently close agreement.

In general, for frequencies of this order it is preferable to employ pentodes; the upper frequency limit with pentodes is above 90 Kc/sec.

Oscillator No. 7.—Low frequency Limit. Using the circuit of fig. 8 but with $C_1 = C_2 = 4 \mu\text{F}$. and with $R_5 = 110,000$ ohms and $R_9 = 100,000$ ohms, an oscillation was obtained with a period of 10 seconds. The wave-form, as estimated from the motion of the pointer of a centre zero microammeter in series with C_2 , was practically sinusoidal. As this seemed to be still above the lowest limit a special circuit was made up to discover this limit.

The circuit used is shown in fig. 11. The feedback circuit employed, that of fig. 3 (c), was used for convenience in wiring. Circuit values

Fig. 11.



were $R_1 = 80,000$ ohms, $R_2 = 25,000$ ohms, $R_3 = 25,000$ ohms, $R_4 = \frac{1}{4}$ megohm and $R_5 = 2$ megohms. R_6 was a 1 megohm variable resistance which was adjusted to control the amplification. C_1 and C_2 were adjusted by switching in $1 \mu\text{F}$. paper condensers.

With $C_1 = C_2 = 1 \mu\text{F}$. the oscillator had a period of 12 seconds, and with $C_1 = 5 \mu\text{F}$., $C_2 = 4 \mu\text{F}$., a period of 50 seconds was obtained; this was the lowest frequency at which this particular circuit would oscillate normally. The wave-form was a reasonably close approximation to a sine wave. At such low frequencies the effect of dielectric polarization is very marked; the capacities of the condensers are very much greater than their nominal values and power losses in the dielectric may be expected to exert an influence.

These very low frequency oscillators provide an excellent demonstration of the action of the oscillator; observation of instrument pointers shows directly the phase relations in the circuit. The effect of increasing amplification is easily seen, the oscillation taking several minutes to build up to a new stable amplitude. The effect of characteristic curvature in

distorting large amplitude oscillations is clearly seen by watching the motion of the anode current meter. It is possible that such an oscillator might be found useful in technical schools for demonstrating phase relations in simple A.C. circuits.

Summary.

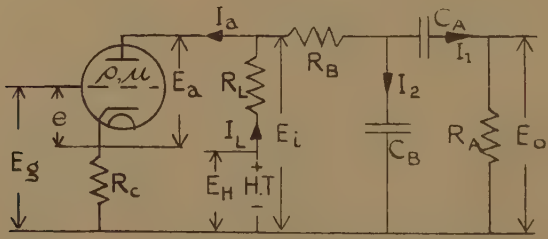
The article describes a push-pull resistance capacity coupled oscillator for frequencies between 0.1 cps. and 80,000 cps. The output wave-form and frequency stability are both very good, and are to a great extent independent of the characteristics of the H.T. and L.T. sources used. Four types of feedback circuit are analysed and compared and the complete oscillator circuit is also analysed. It is shown that a multivibrator, using triode valves and short time constants, can be considered as one form of this type of oscillator. Circuit values are given for oscillators for 500 cps., in which frequency drift is less than one-tenth of one per cent., and details of very low frequency oscillators are given.

APPENDIX I.

Classical Derivation of the Conditions for Oscillation.

As before, one-half of the complete circuit is considered ; the circuit is that shown in fig. 12. Referring to this figure the charge on C_A is Q_1 ; that on C_B is Q_2 .

Fig. 12.



The voltage E_G is the A.C. signal supplied from the other valve, so that

$$\delta E_g = -\delta E_o = -R_A \delta I_1 \dots \dots \dots (i)$$
$$I_L = I_a + I_1 + I_2 \dots \dots \dots (ii)$$
$$E_i = (I_1 + I_2)R_B + I_1 R_A + Q_1/C_A \dots \dots \dots (iii)$$
$$Q_2/C_B = Q_1/C_A + I_1 R_A \dots \dots \dots (iv)$$
$$E_i = E_H - I_L R_L \dots \dots \dots (v)$$
$$\delta I_a = \delta E_a/\rho + \mu \delta e/\rho \dots \dots \dots (vi)$$
assuming straight characteristics ;
$$E_g = e + I_a R_c \dots \dots \dots (vii)$$
$$E_i = E_a + I_a R_c \dots \dots \dots (viii)$$

Let d stand for time differential,

From (vi) and (viii)

$$dI_a = dE_i/\rho + \mu de/\rho - dI_a R_c/\rho.$$

From (vii)

$$dE_g = de + R_c dI_a,$$

therefore

$$dI_a = (dE_i + \mu dE_g)/(\rho + (\mu + 1)R_c). \quad \dots \quad (ix)$$

From (v)

$$dE_i = -R_L dI_L,$$

since E_H is constant H.T.,

$$= -R_L dI_a - R_L dI_1 - R_L dI_2 \text{ from (ii)}$$

$$= \frac{-R_L dE_i}{\rho + (\mu + 1)R_c} - \frac{\mu R_L dE_g}{\rho + (\mu + 1)R_c} - R_L(dI_1 + dI_2) \text{ from (ix).}$$

Therefore

$$dE_i \left[\frac{\rho + R_L + (\mu + 1)R_c}{\rho + (\mu + 1)R_c} \right] = \frac{-\mu R_L dE_g}{\rho + (\mu + 1)R_c} - R_L(dI_1 + dI_2)$$

or

$$dE_i = -A dE_g - R_G(dI_1 + dI_2), \quad \dots \quad (x)$$

$$\text{where } A = \frac{\mu R_L}{\rho + (\mu + 1)R_c + R_L} \quad \text{and} \quad R_G = \frac{R_L[\rho + (\mu + 1)R_c]}{R_L + \rho + (\mu + 1)R_c}.$$

$$\text{From (iv)} \quad I_2/C_B = I_1/C_A + R_A dI_1$$

and

$$dI_2 = C_B dI_1/C_A + C_B R_A d^2 I_1. \quad \dots \quad (xi)$$

From (i)

$$dE_g = -R_A dI_1.$$

Substituting for dI_2 and dE_g in equation (x) gives

$$dE_i = (AR_A - R_G - R_G C_B/C_A) dI_1 - C_B R_G R_A d^2 I_1, \quad \dots \quad (xii)$$

but from (iii)

$$dE_i = R_B dI_2 + R_B dI_1 + R_A dI_1 + I_1/C_A. \quad \dots \quad (xiii)$$

Substituting for dI_2 in this equation and then equating (xii) and (xiii) gives

$$C_A C_B R_A (R_B + R_G) d^2 I_1 + [(R_G + R_B)(C_A + C_B) + C_A R_A - AR_A C_A] dI_1 + I_1 = 0, \quad \dots \quad (xiv)$$

which is oscillatory when

$$A \geq \frac{(R_G + R_B)(C_A + C_B) + C_A R_A}{C_A R_A} \quad \dots \quad (xv)$$

at a frequency of

$$f = \frac{1}{2\pi \sqrt{C_A C_B R_A (R_B + R_G)}} \quad \dots \quad (xvi)$$

Writing

$$C_A R_A = x^2 C_B (R_G + R_B) \quad \text{and} \quad C_B = n C_A,$$

equation (xv) becomes

$$A \geq \frac{n+1+nx^2}{nx^2},$$

which, if x is given the value $\sqrt{1+1/n}$ (to give a maximum value to $\frac{d \tan \phi}{d\omega}$), becomes

$$A \geq 2$$

or

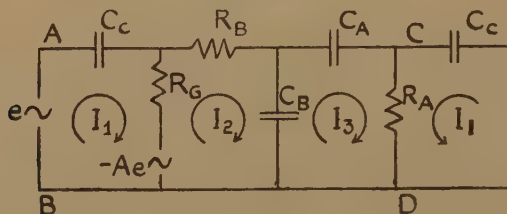
$$\frac{\mu R_L}{R_L + \rho + (\mu + 1)R_c} \geq 2.$$

APPENDIX II.

Equation for Oscillation taking Account of Anode-Grid Capacity.

The circuit may be reduced to that of fig. 13 where the valve and its anode load are replaced by the equivalent generator of fig 5 (c). Referring to the figure the condition for oscillation is that the P.D. between C and

Fig. 13.



D shall be equal in magnitude but opposite in phase to the P.D. between A and B. The current flowing through the anode to grid capacity of the second valve will then be equal in magnitude but opposite in phase to I_1 ; it is shown by a current I_1 circulating in an anti-clockwise direction. The equations connecting the currents and voltages are

$$(R_G - j/\omega C_0)I_1 - R_G I_2 - 0I_3 - (A+1)e = 0$$

$$-R_G I_1 + (R_G + R_B - j/\omega C_B)I_2 + j/\omega C_B I_3 + Ae = 0$$

$$R_A I_1 + j/\omega C_B I_2 + (R_A - j/\omega C_A - j/\omega C_B)I_3 + 0 = 0,$$

where

$$A = \mu R_L / [\rho + R_L + (\mu + 1)R_c].$$

The solutions may be written in determinant form

$$\begin{vmatrix} -R_G & 0 & -(A+1) \\ R_G + R_B - j/\omega C_B & j/\omega C_B & A \\ j/\omega C_B & R_A - j/\omega C_A - j/\omega C_B & 0 \end{vmatrix} \begin{matrix} -I_1 \\ -I_2 \\ -I_3 \end{matrix} = \begin{vmatrix} -(A+1) & R_G - j/\omega C_0 & -R_G \\ A & -R_G & R_G + R_B - j/\omega C_B \\ 0 & R_A & j/\omega C_B \end{vmatrix}$$

$$= \begin{matrix} e \\ \left[\begin{array}{ccc} R_G - j/\omega C_C & -R_G & 0 \\ -R_G & R_G + R_B - j/\omega C_B & j/\omega C_B \\ R_A & j/\omega C_B & R_A - j\omega C_A - j/\omega C_B \end{array} \right] \end{matrix}$$

or
$$\frac{-I_1}{\Delta_1} = \frac{-I_3}{\Delta_3} = \frac{e}{\Delta}.$$

But
$$(I_1 + I_3)R_A = -e,$$

that is
$$(\Delta_1 + \Delta_3)R_A = \Delta,$$

whence by evaluation and elimination two equations are obtained, namely :

$$\frac{AR_A}{\omega C_B} \left(\frac{1}{\omega C_C} - \frac{1}{\omega C_A} \right) = R_A R_B R_G + \frac{R_A + R_B + R_G}{\omega^2 C_B C_C} + \frac{R_A + R_G}{\omega^2 C_A C_B} + \frac{R_B + R_G}{\omega^2 C_A C_C}. \quad (a)$$

and
$$\frac{1}{\omega^2 C_A C_B C_C} = AR_A R_B \left(\frac{1}{C_A} + \frac{1}{C_B} \right) + \frac{R_A R_B + R_A R_G + R_B R_G}{C_A} + \frac{R_A R_B + 4R_A R_G + R_B R_G}{C_B} + \frac{R_A R_G + R_A R_B}{C_C}. \quad (b)$$

The term $R_A R_B R_G$ in equation (a) is negligible if C_A is much greater than C_C . If this term is ignored and the value for A, which is then given by equation (a), is substituted in equation (b) equations (xvii) and (xviii) given in the paper are obtained. Equations (xix) and (xx) are exact expressions, since if R_B is zero the term $R_A R_B R_G$ disappears.

References.

- (1) Terman, Buss, Hewlett and Cahill, Proc. I.R.E. xxvii. p. 649 (1939).
- (2) Scott, Proc. I.R.E. xxvi. p. 226 (1938).
- (3) Aughtie, N. P. L. publication, (E) T.P. p. 247 (1943).
- (4) Van der Pol (1934).

LXXXVI. The "Escalator" Process for the Solution of Lagrangian Frequency Equations.

By J. MORRIS, B.A. and J. W. HEAD, M.A.*

[Received April 7, 1944.]

Summary.

THE "Escalator" process was devised by the authors for the numerical solution of Lagrangian frequency equations, particularly those of high order such as occur in practical problems.

* Communicated by the Director, Royal Aircraft Establishment, Farnborough.

The method is essentially based on the successive introduction or elimination of each of the variables involved by definite self-contained stages. It does not involve the evaluation of determinants; neither does it involve iteration. The only successive approximation device used is the Newtonian method for approximating to the roots of an equation.

An important feature of the process consists of powerful checks at each stage to ensure accuracy and if necessary to adjust it.

With these checks it appears possible to maintain an accuracy which falls off roughly at the rate of one significant figure per three stages.

From the experience gained by the application of the Escalator process to two twelfth-order equations, the method has been found to be practicable and fairly easy to work.

The authors have endeavoured to keep the analysis on simple algebraic lines.

CONTENTS.

	Page
Introduction	736
1. Orthogonal properties of modes	738
2. The "bordered" equations	740
3. The "Escalator"	741
4. Checks	741
5. Separation of roots	741
6. The "Escalator in Reverse"	741
7. Notation.....	743
8. Canonical equations	743
9. Special canonical form	746
10. Equal and/or repeated roots	746
11. Equal and/or complex roots for the unsymmetrical canonical case	749
12. Reduction of a root to zero	751
13. Latent root equations arising from solution of linear simultaneous equations by iteration processes ...	752
14. The solution of canonical Lagrangian frequency equations by means of linear simultaneous equations	754
References	756
Appendix—Numerical example of application of process	756

Introduction.

The orthodox solution of Lagrangian frequency equations involves the expansion into polynomial form of the characteristic determinantal equation in the "latent" roots and, when these have been found, the relative "modes" are ascertained from determinantal expressions of the kind arising in the solution of linear simultaneous equations. Where the frequency equations are of a comparatively low order, say below the sixth order, the numerical computation involved is not unduly laborious but for higher orders it becomes increasingly arduous if results accurate to four or more significant figures are required. Moreover, there are no

ready means of becoming aware at any particular stage whether a mistake has been made in the arithmetic ; it is only when checking the values in the original equations that serious errors are revealed.

An interesting and useful contribution to the problem was made by Messrs. W. J. Duncan and A. R. Collar in 1934, when they produced their matrix iteration solution⁽¹⁾. Their method was devised for the solution of equations in which the latent root occurs only along the leading diagonal, in which case an iteration process was found to be readily applicable. First, the highest root is found together with its associated modes. Then by using an artifice, based on the orthogonal properties of the modes, the equations are reduced by one degree by means analogous to the elimination of one of the variables in a set of linear simultaneous equations. The highest root is thereby separated from the original equations and the iteration process is then applied to find the next highest root and its associated modes and so on. But it has been found by experience that for equations of the ninth and higher orders the iteration involves a great strain. Assuming, for example, that we require the first seven roots and their associated modes of an equation of high order, accurate to six significant figures, we should be obliged to calculate the first root and its associated modes to twelve significant figures owing to the inevitable falling off of accuracy with successive eliminations. This drop in accuracy in the iteration process has been found by experience to be roughly at the rate of one significant figure per elimination. In such circumstances the arithmetic becomes exceedingly laborious.

In consequence of these difficulties and other considerations the authors devised the "Escalator" process when they were confronted with a practical problem in vibration which involved the numerical solution of a twelfth-order Lagrangian frequency equation.

The method is essentially based on the successive introduction or elimination of each of the variables involved by definite self-contained stages in which the roots and modes of say a fourth-order equation are obtained in terms of those of the preceding third-order equation and *vice versa*.

The escalator process does not involve the evaluation of determinants ; neither does it involve iteration. The only successive approximation device used is the Newtonian method for approximating to the roots of an equation. A paramount feature of the escalator process consists of powerful checks applicable at each stage to ensure accuracy and if necessary to adjust it. This makes it unnecessary to go through the laborious process of checking results in the original equations at the final or at any intermediate stage.

With these checks it appears possible to maintain an accuracy which falls off roughly one significant figure per three stages. Thus, in two twelfth-order cases which have been worked out by the process, it has been found that keeping to eight significant figures throughout the

calculations, which were done by calculating machine, detailed results appeared to be accurate to some five significant figures.

To assess the possibilities of the escalator process, suppose we have solved a set of equations of high order and wish to know what happens if certain alterations in the elements of a particular "row" and "column" are made. By the use of the escalator in reverse we can find the results for the equations from which the row and column in question have been eliminated. We can then re-escalate with the new row and column, thereby making the correction with remarkable facility.

A brief account of the escalator process⁽²⁾ was published by the authors in November, 1942. In the treatment there adopted, as in the present paper, it was decided to keep the analysis on simple algebraic lines. It may happen that those well versed in the matrix notation will find it easier to follow the process by working it out by that notation, but the authors set out to cater for those who, like the authors themselves, have minds which are not readily attuned to the matrix notation.

Much attention has been given in recent years to moment distribution⁽³⁾ and relaxation⁽⁴⁾ methods because of the difficulty of solving linear simultaneous and Lagrangian frequency equations of high order which occur in practical problems. But these methods are merely physical applications of algebraic iteration processes. In practical problems these processes are usually convergent, but it has been thought worth while to give a brief account of the conditions of convergence of the iteration process applied to linear simultaneous equations, which was discussed elsewhere by one of the authors⁽⁵⁾.

In general the authors consider that it is simpler to reduce the practical problem where applicable to sets of linear "normal" equations or Lagrangian frequency equations, as the case may be, and then proceed to solve these by iteration or escalation respectively.

1. Orthogonal Properties of Modes.

Consider the following fourth-order Lagrangian frequency equations:—

$$(a_{11}-b_{11}\lambda)x+(a_{12}-b_{12}\lambda)y+(a_{13}-b_{13}\lambda)z+(a_{14}-b_{14}\lambda)u=0, \quad (1)$$

$$(a_{12}-b_{12}\lambda)x+(a_{22}-b_{22}\lambda)y+(a_{23}-b_{23}\lambda)z+(a_{24}-b_{24}\lambda)u=0, \quad (2)$$

$$(a_{13}-b_{13}\lambda)x+(a_{23}-b_{23}\lambda)y+(a_{33}-b_{33}\lambda)z+(a_{34}-b_{34}\lambda)u=0, \quad (3)$$

$$(a_{14}-b_{14}\lambda)x+(a_{24}-b_{24}\lambda)y+(a_{34}-b_{34}\lambda)z+(a_{44}-b_{44}\lambda)u=0, \quad (4)$$

where the coefficients a_{rs} , b_{rs} are real, λ is a "latent" root and x , y , z , u its associated "modes."

If either the a 's or b 's are coefficients of a positive homogeneous quadratic function, then the latent roots will be real. If both the a 's and b 's are such coefficients then the latent roots will also be positive.

Let λ_r , x_r , y_r , z_r ($r=1, 2, 3$) be a solution to the third-order equations (1), (2), (3), with $u=0$, and let

$$x_r' = b_{11}x_r + b_{12}y_r + b_{13}z_r = (a_{11}x_r + a_{12}y_r + a_{13}z_r)/\lambda_r, \quad (5)$$

$$y_r' = b_{12}x_r + b_{22}y_r + b_{23}z_r = (a_{12}x_r + a_{22}y_r + a_{23}z_r)/\lambda_r, \quad (6)$$

$$z_r' = b_{13}x_r + b_{23}y_r + b_{33}z_r = (a_{13}x_r + a_{23}y_r + a_{33}z_r)/\lambda_r. \quad (7)$$

We shall call x_r', y_r', z_r' "transposed" modes.

Multiply the relations (5), (6), (7) respectively by x_s, y_s, z_s , where s differs from r , and add.

$$\text{We obtain} \quad x_r'x_s + y_r'y_s + z_r'z_s = B_{rs} = A_{rs}/\lambda_r, \quad (8)$$

$$\text{where} \quad A_{rs} = a_{11}x_rx_s + a_{22}y_ry_s + a_{33}z_rz_s + a_{12}(x_sy_r + x_ry_s) \\ + a_{13}(x_zy_r + x_rz_s) + a_{23}(y_sz_r + y_rz_s), \quad (9)$$

and B_{rs} is a similar expression with b in place of a .

In the same way we find that

$$x_rx_s' + y_ry_s' + z_rz_s' = B_{sr} = A_{sr}/\lambda_s, \quad (10)$$

in which $A_{sr} = A_{rs}$; $B_{sr} = B_{rs}$.

Thus, if λ_s is not equal to λ_r it follows that

$$A_{sr} = A_{rs} = 0; \quad B_{sr} = B_{rs} = 0. \quad (11)$$

Hence, in general, when r and s are unequal

$$x_r'x_s + y_r'y_s + z_r'z_s = 0. \quad (12)$$

Now the frequency equations only give the ratios of the modes; we may, therefore, without loss of generality make

$$x_r'x_r + y_r'y_r + z_r'z_r = 1. \quad (13)$$

We shall call modes subject to this condition "rectified" modes.

Rectified modes will have properties analogous to those of the direction cosines of a rectangular system of coordinate axes. This may be readily proved for the third-order equations as follows:—

Let

$$X' = x_1X + x_2Y + x_3Z, \quad Y' = y_1X + y_2Y + y_3Z, \quad Z' = z_1X + z_2Y + z_3Z,$$

where x_r, y_r, z_r are rectified modes associated with the root λ_r ($r=1, 2, 3$) and X, Y, Z are any quantities whatever. We have at once

$$x_1'X' + y_1'Y' + z_1'Z' = X, \quad x_2'X' + y_2'Y' + z_2'Z' = Y, \\ x_3'X' + y_3'Y' + z_3'Z' = Z,$$

where x_r', y_r', z_r' ($r=1, 2, 3$) are the transposed rectified modes associated with x_r, y_r, z_r , as specified by (5), (6) and (7).

Substituting these values in the expression for X' we obtain the identity

$$X' \equiv (x_1x_1' + x_2x_2' + x_3x_3')X' + (x_1y_1' + x_2y_2' + x_3y_3')Y' \\ + (x_1z_1' + x_2z_2' + x_3z_3')Z'. \quad (14)$$

Hence it follows that

$$x_1x_1' + x_2x_2' + x_3x_3' = 1, \quad (15)$$

$$x_1y_1' + x_2y_2' + x_3y_3' = 0, \quad (16)$$

$$x_1z_1' + x_2z_2' + x_3z_3' = 0. \quad (17)$$

Similar relations follow from the expressions for Y' , Z' .

Although in the foregoing we have confined our attention to third-order equations, it is evident that the propositions apply to equations of any order.

2. The "Bordered" Equations.

Suppose now we consider the fourth-order equations (1), (2), (3) (4), obtained by bordering the third-order equations (1), (2), (3), with $u=0$.

Let $\lambda_r (r=1, 2, 3)$ be a latent root of the third-order equations and let x_r, y_r, z_r be its associated rectified modes with x'_r, y'_r, z'_r the corresponding transposed rectified modes.

Multiply equations (1), (2), (3) respectively by x_r, y_r, z_r and add. We obtain

$$(\lambda_r - \lambda)(x'_r x + y'_r y + z'_r z) + P_r u = 0, \quad . \quad . \quad . \quad (18)$$

where $P_r = (a_{14} - b_{14}\lambda)x_r + (a_{24} - b_{24}\lambda)y_r + (a_{34} - b_{34}\lambda)z_r. \quad . \quad . \quad (19)$

Let $P = (a_{14} - b_{14}\lambda)x + (a_{24} - b_{24}\lambda)y + (a_{34} - b_{34}\lambda)z, \quad . \quad . \quad (20)$

which in accordance with equation (4) is the same as

$$P = -(a_{44} - b_{44}\lambda)u. \quad . \quad . \quad . \quad (21)$$

Now having regard to the "rectangular" properties of the rectified modes it follows that

$$\sum_{r=1}^{r=3} P_r (x'_r x + y'_r y + z'_r z) = P. \quad . \quad . \quad . \quad (22)$$

Hence the latent root equation for the fourth-order equations (1), (2), (3), (4) may be written

$$\sum_{r=1}^{r=3} P_r^2 / (\lambda_r - \lambda) = a_{44} - b_{44}\lambda, \quad . \quad . \quad . \quad (23)$$

where $\lambda_r (r=1, 2, 3)$ is a latent root of the third-order equations (1), (2), (3), with $u=0$, and P_r is as defined by (19).

Having found the λ roots of (23) we obtain the corresponding modes for the fourth-order equations as follows:—

From equations of the form (18), and having regard to the rectangular properties of the rectified modes, it is easily seen that

$$x/u = - \sum_{r=1}^{r=3} P_r x'_r / (\lambda_r - \lambda), \quad . \quad . \quad . \quad (24)$$

$$y/u = - \sum_{r=1}^{r=3} P_r y'_r / (\lambda_r - \lambda), \quad . \quad . \quad . \quad (25)$$

$$z/u = - \sum_{r=1}^{r=3} P_r z'_r / (\lambda_r - \lambda); \quad . \quad . \quad . \quad (26)$$

and from these it follows that

$$x'/u = b_{14} - \sum_{r=1}^{r=3} P_r x'_r / (\lambda_r - \lambda), \quad . \quad . \quad . \quad (27)$$

$$y'/u = b_{24} - \sum_{r=1}^{r=3} P_r y'_r / (\lambda_r - \lambda). \quad . \quad . \quad . \quad (28)$$

$$z'/u = b_{34} - \sum_{r=1}^{r=3} P_r' z_r / (\lambda_r - \lambda). \quad (29)$$

We notice that

$$\sum_{r=1}^{r=3} P_r x_r' = (a_{14} - b_{14}\lambda), \quad \sum_{r=1}^{r=3} P_r y_r' = (a_{24} - b_{24}\lambda), \quad \sum_{r=1}^{r=3} P_r z_r' = (a_{34} - b_{34}\lambda). \quad (30)$$

It also follows that

$$(xx' + yy' + zz' + uu')/u^2 = 2u'/u - b_{44} + \sum_{r=1}^{r=3} P_r^2 / (\lambda_r - \lambda)^2 \quad (31)$$

If we write equation (23) as

$$f(\lambda) = -a_{44} + b_{44}\lambda + \sum P_r^2 / (\lambda_r - \lambda) = 0, \quad (32)$$

we have
$$f'(\lambda) = b_{44} + \sum P_r^2 / (\lambda_r - \lambda)^2 + 2 \sum P_r \frac{dP_r}{d\lambda} / (\lambda_r - \lambda), \quad (33)$$

and since
$$\frac{dP_r}{d\lambda} = -(b_{14}x_r + b_{24}y_r + b_{34}z_r), \quad (34)$$

it may be shown that the right-hand side of (33) is identical with the right-hand side of (31).

Hence the rectified value of u^2 is given by the relation $1/u^2 = f'(\lambda)$ where λ is a root of $f(\lambda) = 0$.

3. The "Escalator."

We have proved that, given the solution of a set of Lagrangian frequency equations of any order, we can write down the latent root equation for a succeeding order, derived by bordering the original equations by an additional row and column; and in addition we have obtained expressions for the modes of the succeeding order. We shall call the process "escalation" and the particular form of the latent root equation (23) the "escalator" equation. We notice that the escalator equation is in the form of simple partial fractions convenient for the numerical evaluation of the latent roots.

4. Checks.

The relations (30), (31) and (33) give useful and powerful checks on the arithmetic; the first for the P 's and the other two for the rectification of the modes.

5. Separation of Roots.

It follows from the form of the escalator equation that the latent roots of any order are separated by those for the preceding order.

6. The "Escalator in Reverse."

So far the escalator has been demonstrated for ascending orders. It will now be shown that it can also be used for descending orders. Thus, let $\lambda_r (r=1, 2, 3, 4)$ be a latent root of the fourth-order equations (1), (2), (3), (4) and let x_r, y_r, z_r, u_r be its associated rectified modes and x_r', y_r', z_r', u_r' the corresponding rectified transposed modes. To find the

solution, in terms of these, of the third-order equations (1), (2), (3), with $u=0$.

Multiply the third-order equations respectively by x_r , y_r , z_r and add. We obtain

$$(\lambda_r - \lambda)(x_r'x + y_r'y + z_r'z) = Pu_r, \quad (35)$$

$$\text{where} \quad P = (a_{14} - b_{14}\lambda)x + (a_{24} - b_{24}\lambda)y + (a_{34} - b_{34}\lambda)z. \quad (36)$$

$$\text{Hence} \quad \sum_{r=1}^{r=4} u_r(x_r'x + y_r'y + z_r'z) = P \sum_{r=1}^{r=4} u_r^2/(\lambda_r - \lambda). \quad (37)$$

But in virtue of the rectangular properties of the rectified modes $\Sigma u_r x_r' = 0$, $\Sigma u_r y_r' = 0$, $\Sigma u_r z_r' = 0$; hence we derive from (37) the "escalator in reverse" equation for the latent roots of the third-order equation, viz.:

$$\sum_{r=1}^{r=4} u_r^2/(\lambda_r - \lambda) = 0. \quad (38)$$

For the modes pertaining to a root λ of (38) we find

$$x/P = \Sigma u_r x_r/(\lambda_r - \lambda), \quad y/P = \Sigma y_r u_r/(\lambda_r - \lambda), \quad z/P = \Sigma z_r u_r/(\lambda_r - \lambda), \quad (39)$$

where P is as defined by (36) and the suffixed quantities refer to the fourth-order equations from which we are deriving the solution of the third-order equations.

Again, for the third-order equations

$$x' = b_{11}x + b_{12}y + b_{13}z; \quad (40)$$

$$\therefore x'/P = \Sigma (x_r' - b_{14}u_r)u_r/(\lambda_r - \lambda), \quad (41)$$

$$\text{or having regard to (38)} \quad x'/P = \Sigma x_r' u_r/(\lambda_r - \lambda). \quad (42)$$

$$\text{Similarly} \quad y'/P = \Sigma y_r' u_r/(\lambda_r - \lambda), \quad z'/P = \Sigma z_r' u_r/(\lambda_r - \lambda). \quad (43)$$

For rectification of these modes we have

$$\begin{aligned} (xx' + yy' + zz')/P^2 &= \Sigma x_r u_r/(\lambda_r - \lambda) \times \Sigma x_r' u_r/(\lambda_r - \lambda) \\ &\quad + \Sigma y_r u_r/(\lambda_r - \lambda) \times \Sigma y_r' u_r/(\lambda_r - \lambda) \\ &\quad + \Sigma z_r u_r/(\lambda_r - \lambda) \times \Sigma z_r' u_r/(\lambda_r - \lambda). \end{aligned} \quad (44)$$

$$\text{But R.H.S. of (44)} = \Sigma (1 - u_r u_r') u_r^2/(\lambda_r - \lambda)^2$$

$$\begin{aligned} &\quad - \Sigma (u_r u_s' + u_r' u_s) u_r u_s/(\lambda_r - \lambda)(\lambda_s - \lambda) \\ &= \Sigma u_r^2/(\lambda_r - \lambda)^2 - \Sigma u_r^2/(\lambda_r - \lambda) \times \Sigma u_r u_r'/(\lambda_r - \lambda). \end{aligned} \quad (45)$$

Hence, in virtue of (38), the rectified value of P for the root λ of (38) will be given by

$$1/P^2 = \Sigma u_r^2/(\lambda_r - \lambda)^2. \quad (46)$$

Expressing the escalator in reverse equation (38) as $f(\lambda) = 0$, we find that the rectified value of P is given by the relation $1/P^2 = f'(\lambda)$, where λ is a root of $f(\lambda) = 0$. The rectified values of x , y , z , x' , y' , z' are then given by (39), (41), (43) with the value of P given by (46).

7. Notation.

We can if we choose adopt the notation λ_{sr} to denote the r th λ root of the s th order equation, with corresponding designation for its associated modes. P_{sr} may be used to denote the P for the s th order derived from the elements of the s th order and the modal values applicable to the r th λ root of the preceding $(s-1)$ th order.

Thus the third-order equations would be written

$$(a_{11}-b_{11}\lambda_3)x_3+(a_{12}-b_{12}\lambda_3)y_3+(a_{13}-b_{13}\lambda_3)z_3=0, \quad . \quad . \quad (47)$$

$$(a_{12}-b_{12}\lambda_3)x_3+(a_{22}-b_{22}\lambda_3)y_3+(a_{23}-b_{23}\lambda_3)z_3=0, \quad . \quad . \quad (48)$$

$$(a_{13}-b_{13}\lambda_3)x_3+(a_{23}-b_{23}\lambda_3)y_3+(a_{33}-b_{33}\lambda_3)z_3=0. \quad . \quad . \quad (49)$$

The solution of these equations would be expressed as λ_{3r} , x_{3r} , y_{3r} , z_{3r} , x_{3r}' , y_{3r}' , z_{3r}' ($r=1, 2, 3$). For the fourth-order we should have

$$P_{4r}=(a_{14}-b_{14}\lambda_4)x_{3r}+(a_{24}-b_{24}\lambda_4)y_{3r}+(a_{34}-b_{34}\lambda_4)z_{3r}; \quad . \quad (50)$$

the escalator equation for the fourth-order would be

$$\Sigma P_{4r}^2/(\lambda_{3r}-\lambda_4)=a_{44}-b_{44}\lambda_4, \quad . \quad . \quad . \quad (51)$$

and the relative modes for the root λ_4 ,

$$x_4/u_4=-\sum_{r=1}^{r=3} P_{4r}x_{3r}/(\lambda_{3r}-\lambda_4), \quad . \quad . \quad . \quad (52)$$

and so on.

For the escalator in reverse, say from the fourth-order to the third-order, we should have λ_{4r} , x_{4r} , y_{4r} , z_{4r} , u_{4r} , x_{4r}' , y_{4r}' , z_{4r}' , u_{4r}' pertaining to the fourth-order and the derived third-order escalator equation expressed as

$$\Sigma u_{4r}^2/(\lambda_{4r}-\lambda_3)=0. \quad . \quad . \quad . \quad (53)$$

The relative modes for the root λ_3 of the third-order would be

$$x_3/P_3=\Sigma x_{4r}u_{4r}/(\lambda_{4r}-\lambda_3), \text{ etc.}, \quad . \quad . \quad . \quad (54)$$

where $P_3=(a_{14}-b_{14}\lambda_3)x_3+(a_{24}-b_{24}\lambda_3)y_3+(a_{34}-b_{34}\lambda_3)z_3. \quad . \quad . \quad (55)$

8. Canonical Equations.

Equations in which the latent root occurs only along the leading diagonal are called for convenience canonical equations. Let us consider the following unsymmetrical set of fourth-order equations, viz. :—

$$\lambda x=a_{11}x+a_{21}y+a_{31}z+a_{41}u, \quad . \quad . \quad . \quad (56)$$

$$\lambda y=a_{12}x+a_{22}y+a_{32}z+a_{42}u, \quad . \quad . \quad . \quad (57)$$

$$\lambda z=a_{13}x+a_{23}y+a_{33}z+a_{43}u, \quad . \quad . \quad . \quad (58)$$

$$\lambda u=a_{14}x+a_{24}y+a_{34}z+a_{44}u. \quad . \quad . \quad . \quad (59)$$

We first form the “transposed” set of equations, viz. :

$$\lambda x'=a_{11}x'+a_{12}y'+a_{13}z'+a_{14}u', \quad . \quad . \quad . \quad (60)$$

$$\lambda y'=a_{21}x'+a_{22}y'+a_{23}z'+a_{24}u', \quad . \quad . \quad . \quad (61)$$

$$\lambda z' = a_{31}x' + a_{32}y' + a_{33}z' + a_{34}u', \quad . \quad . \quad . \quad (62)$$

$$\lambda u' = a_{41}x' + a_{42}y' + a_{43}z' + a_{44}u'. \quad . \quad . \quad . \quad (63)$$

The transposed equations will clearly have the same latent roots as the original equations. Let λ_r ($r=1, 2, 3$) be a latent root of the third-order equations (56), (57), (58), with $u=0$, and let the relative modes be x_r, y_r, z_r and the relative transposed modes x'_r, y'_r, z'_r .

Multiply these third-order equations respectively by x'_r, y'_r, z'_r and add. We obtain

$$(\lambda_r - \lambda)(x'_r x + y'_r y + z'_r z) = 0, \quad . \quad . \quad . \quad (64)$$

from which it follows that

$$x'_r x + y'_r y + z'_r z = 0, \quad . \quad . \quad . \quad (65)$$

provided r is not equal to s .

The modal property (65) is of the same form as (12) for the Lagrangian frequency equations. It follows that rectified modal properties similar to those for the Lagrangian case apply to the unsymmetrical canonical case. In this connection it is to be noticed that if the Lagrangian equations (1), (2), (3), with $u=0$, are treated as linear simultaneous equations in $\lambda x, \lambda y, \lambda z$, we shall obtain an unsymmetrical set of canonical equations of the type (56), (57), (58), with $u=0$; but it may be shown that x', y', z' , as defined by (5), (6), (7) satisfy the corresponding transposed equations of the type (60), (61), (62), with $u=0$.

Suppose, now, we consider the fourth-order equations (56), (57), (58), (59) derived by bordering the third-order equations. Regarding λ_r as a root of the third-order equations and x_r, y_r, z_r as its associated rectified modes with x'_r, y'_r, z'_r the corresponding rectified transposed modes, we multiply (56), (57), (58) respectively by x'_r, y'_r, z'_r and add. We obtain

$$(\lambda_r - \lambda)(x'_r x + y'_r y + z'_r z) + P_r u = 0, \quad . \quad . \quad . \quad (66)$$

where

$$P_r = a_{41}x'_r + a_{42}y'_r + a_{43}z'_r. \quad . \quad . \quad . \quad (67)$$

Let

$$P_r = a_{14}x_r + a_{24}y_r + a_{34}z_r; \quad . \quad . \quad . \quad (68)$$

then since $x_r, y_r, z_r, x'_r, y'_r, z'_r$ are rectified modes we have

$$\sum_{r=1}^{r=3} P_r (x'_r x + y'_r y + z'_r z) = P. \quad . \quad . \quad . \quad (69)$$

where

$$P = a_{14}x + a_{24}y + a_{34}z, \quad . \quad . \quad . \quad (70)$$

or

$$P = -(a_{44} - \lambda)u.$$

It follows, therefore, that

$$\sum_{r=1}^{r=3} P_r P_r \lambda_r - P/(\lambda) = a_{44} - \lambda, \quad . \quad . \quad . \quad (71)$$

will be the escalator latent root equation for the fourth-order, and that the relative modes will be given by

$$\begin{aligned} x/u &= -\Sigma P'_r x'_r / (\lambda_r - \lambda), & y/u &= -\Sigma P'_r y'_r / (\lambda_r - \lambda), \\ z/u &= -\Sigma P'_r z'_r / (\lambda_r - \lambda); & x'/u' &= -\Sigma P_r x'_r / (\lambda_r - \lambda), \\ y'/u' &= -\Sigma P_r y'_r / (\lambda_r - \lambda), & z'/u' &= -\Sigma P_r z'_r / (\lambda_r - \lambda). \end{aligned} \quad . \quad . \quad (72)$$

Also, it may be shown that

$$(xx' + yy' + zz' + uu')/uu' = 1 + \Sigma P_r P_r' / (\lambda_r - \lambda)^2. \quad (73)$$

For rectification of these modes we may, without loss of generality, make $u' = \pm u$, and thus the rectified value of u will be given by

$$1/u^2 = \pm [1 + \Sigma P_r P_r' / (\lambda_r - \lambda)^2]. \quad (74)$$

The other rectified modes will follow from (72).

Again, we notice that if

$$f(\lambda) = -a_{44} + \lambda + \Sigma P_r P_r' / (\lambda_r - \lambda) = 0, \quad (75)$$

then

$$1/u^2 = \pm f'(\lambda), \quad (76)$$

in which we choose the sign which makes u^2 positive.

For the escalator in reverse in this case let λ_r ($r=1, 2, 3, 4$) be a latent root of the fourth-order equations (56), (57), (58), (59); let x_r, y_r, z_r, u_r be its rectified modes and x_r', y_r', z_r', u_r' the corresponding rectified transposed modes. Multiply the third-order equations (56), (57), (58) with $u=0$, respectively by x_r', y_r', z_r' and add. We obtain

$$(\lambda_r - \lambda)(x_r'x + y_r'y + z_r'z) = Pu_r', \quad (77)$$

where

$$P = a_{14}x + a_{24}y + a_{34}z.$$

From the form of (77) and having regard to the rectangular properties of the rectified modes, we derive the escalator equation for the third-order equations, viz.:

$$\sum_{r=1}^{r=4} u_r u_r' / (\lambda_r - \lambda) = 0. \quad (78)$$

For the modes associated with a root λ of (78) we find

$$\begin{aligned} x/P &= \Sigma x_r u_r' / (\lambda_r - \lambda), & y/P &= \Sigma y_r u_r' / (\lambda_r - \lambda), \\ z/P &= \Sigma z_r u_r' / (\lambda_r - \lambda); \end{aligned}$$

and similarly, for the transposed modes,

$$\begin{aligned} x'/P' &= \Sigma x_r' u_r / (\lambda_r - \lambda), & y'/P' &= \Sigma y_r' u_r / (\lambda_r - \lambda), \\ z'/P' &= \Sigma z_r' u_r / (\lambda_r - \lambda), \end{aligned} \quad (79)$$

where

$$P' = a_{41}x' + a_{42}y' + a_{43}z'.$$

For rectification of the modes it may be shown that

$$(xx' + yy' + zz')/PP' = \Sigma u_r u_r' / (\lambda_r - \lambda)^2. \quad (80)$$

We may, without loss of generality, take PP' to be $\pm P^2$, the sign being chosen so as to make P^2 positive. We thus have for the rectified value of P

$$1/P^2 = \pm \Sigma u_r u_r' / (\lambda_r - \lambda)^2. \quad (81)$$

We notice that in the forward escalation for the unsymmetrical canonical case, say, in passing from the third to the fourth-order

$$\Sigma P_r P_r' = a_{41}a_{14} + a_{42}a_{24} + a_{43}a_{34}; \quad (82)$$

$$\begin{aligned} \Sigma P_r' x_r &= a_{41}, & \Sigma P_r' y_r &= a_{42}, & \Sigma P_r' z_r &= a_{43}, \\ \Sigma P_r x_r' &= a_{14}, & \Sigma P_r y_r' &= a_{24}, & \Sigma P_r z_r' &= a_{34}. \end{aligned} \quad (83)$$

We also notice that $\Sigma\lambda_r = \Sigma a_{rr}$, where $\Sigma\lambda_r$ is the sum of the λ roots of any particular order and Σa_{rr} is the sum of the leading diagonal coefficients for that order.

9. Special Canonical Form.

Suppose the latent root equations we have to solve are of the form

$$(m_1 p_{11} - \lambda)Y_1 + m_2 p_{12}Y_2 + m_3 p_{13}Y_3 = 0, \quad . \quad . \quad . \quad (84)$$

$$m_1 p_{12}Y_1 + (m_2 p_{22} - \lambda)Y_2 + m_3 p_{23}Y_3 = 0, \quad . \quad . \quad . \quad (85)$$

$$m_1 p_{13}Y_1 + m_2 p_{23}Y_2 + (m_3 p_{33} - \lambda)Y_3 = 0, \quad . \quad . \quad . \quad (86)$$

in which λ and the Y 's are variables.

Let $\sqrt{m_1}Y_1 = x$, $\sqrt{m_2}Y_2 = y$, $\sqrt{m_3}Y_3 = z$; $m_1 p_{11} = a_{11}$, $m_2 p_{22} = a_{22}$,
 $m_3 p_{33} = a_{33}$, $\sqrt{m_1 m_2} p_{12} = a_{12}$, $\sqrt{m_1 m_3} p_{13} = a_{13}$, $\sqrt{m_2 m_3} p_{23} = a_{23}$;

then equations (84), (85), (86) may be written

$$(a_{11} - \lambda)x + a_{12}y + a_{13}z = 0, \quad . \quad . \quad . \quad (87)$$

$$a_{12}x + (a_{22} - \lambda)y + a_{23}z = 0, \quad . \quad . \quad . \quad (88)$$

$$a_{13}x + a_{23}y + (a_{33} - \lambda)z = 0. \quad . \quad . \quad . \quad (89)$$

Symmetrical canonical equations of this type we shall designate as of "special canonical form." In such case the transposed modes x' , y' , z' are identical with the ordinary x , y , z modes respectively.

10. Equal and/or Repeated Roots in Escalator Process.

In the case of Lagrangian frequency equations the coefficients are usually coefficients of positive homogeneous quadratic functions, in which circumstances the latent roots will be real and positive. Also there will be separation of roots by successive orders. There may, however, be equal and/or repeated roots. We shall therefore discuss such roots and how they affect the escalator process. Consider, for example, the fourth-order equations (1), (2), (3), (4). With the notation of § 7, let λ_{21} , λ_{22} be the roots of the second-order equations (1), (2), with $z=0$ and $u=0$, and let x_{21} , y_{21} , x_{21}' , y_{21}' ; x_{22} , y_{22} , x_{22}' , y_{22}' be the respective rectified modes.

We have for the third-order equations (1), (2), (3), with $u=0$,

$$(\lambda_{21} - \lambda_3)(x_3 x_{21}' + y_3 y_{21}') + P_{31} z_3 = 0, \quad . \quad . \quad . \quad (90)$$

$$(\lambda_{22} - \lambda_3)(x_3 x_{22}' + y_3 y_{22}') + P_{32} z_3 = 0. \quad . \quad . \quad . \quad (91)$$

If $\lambda_{31} = \lambda_{21}$ is a root of the third-order equations then these equations will be satisfied for this root with modes

$$x_{31} = x_{21}, \quad y_{31} = y_{21}, \quad z_{31} = 0,$$

$$x_{31}' = x_{21}', \quad y_{31}' = y_{21}', \quad z_{31}' = b_{13}x_{21} + b_{23}y_{21};$$

and in addition $(P_{31})_{\lambda_{31}}$ must be zero, i. e.

$$(a_{13} - b_{13}\lambda_{21})x_{21} + (a_{23} - b_{23}\lambda_{21})y_{21} = 0. \quad . \quad . \quad . \quad (92)$$

Now

$$P_{31} = (a_{13} - b_{13}\lambda_3)x_{21} + (a_{23} - b_{23}\lambda_3)y_{21},$$

or, in the case when (92) holds good,

$$P_{31}=(\lambda_{21}-\lambda_3)(b_{13}x_{21}+b_{23}y_{21}). \quad (93)$$

Equation (90) may now be written

$$x_3x_{21}'+y_3y_{21}'+z_3(b_{13}x_{21}+b_{23}y_{21})=0, \quad (94)$$

in which the root λ_{21} has been eliminated. From (94), (91) and the last of the third-order equations we derive the escalator equation in λ for the other two roots λ_{32} , λ_{33} of the third-order equations, viz.:

$$(b_{13}x_{21}+b_{23}y_{21})^2(\lambda_{21}-\lambda_3)+P_{32}^2/(\lambda_{22}-\lambda_3)=a_{33}-b_{33}\lambda_3, \quad (95)$$

the relative modes for which will be given by

$$x_3/z_3=-[(b_{13}x_{21}+b_{23}y_{21})x_{21}+P_{32}x_{22}/(\lambda_{22}-\lambda_3)], \quad (96)$$

$$y_3/z_3=-[(b_{13}x_{21}+b_{23}y_{21})y_{21}+P_{32}y_{22}/(\lambda_{22}-\lambda_3)]; \quad (97)$$

$$x_3'/z_3=b_{13}-[(b_{13}x_{21}+b_{23}y_{21})x_{21}'+P_{32}x_{22}'/(\lambda_{22}-\lambda_3)], \quad (98)$$

$$y_3'/z_3=b_{23}-[(b_{13}x_{21}+b_{23}y_{21})y_{21}'+P_{32}y_{22}'/(\lambda_{22}-\lambda_3)]. \quad (99)$$

Suppose that $\lambda_{32}=\lambda_{31}=\lambda_{21}$ is also a root of the escalator equation (95) and let its rectified modes as given by (96), (97), (98), (99) be x_{32} , y_{32} , z_{32} , x_{32}' , y_{32}' , z_{32}' .

We thus have two equal roots for the third-order equations but with distinct rectangular modes. These modes are, however, subject to a certain indefiniteness in that for either of the equal roots the equations will be satisfied by any linear combination of the two sets of distinct modes. Thus we may have the relative modal combinations

$$\begin{aligned} &x_{31}+p_1x_{32}, \quad y_{31}+p_1y_{32}, \quad z_{31}+p_1z_{32}, \\ &x_{31}'+p_1x_{32}', \quad y_{31}'+p_1y_{32}', \quad z_{31}'+p_1z_{32}'; \\ &x_{31}+p_2x_{32}, \quad y_{31}+p_2y_{32}, \quad z_{31}+p_2z_{32}, \\ &x_{31}'+p_2x_{32}', \quad y_{31}'+p_2y_{32}', \quad z_{31}'+p_2z_{32}'; \end{aligned}$$

which modes will be orthogonal to the remaining modes and which, in addition, will be mutually orthogonal provided

$$1+p_1p_2=0 \quad (100)$$

For rectification the two above sets of modes require division by $\sqrt{1+p_1^2}$, $\sqrt{1+p_2^2}$ respectively.

Thus, if p_1 is arbitrary, the appropriate value of p_2 to comply with orthogonality will be given by (100).

Let λ_{33} be the remaining root of (95) and let x_{33} , y_{33} , z_{33} , x_{33}' , y_{33}' , z_{33}' be its associated rectified modes.

Consider next the fourth-order equations for which we shall have in the usual way

$$(\lambda_{31}-\lambda_4)(x_4x_{31}'+y_4y_{31}'+z_4z_{31}')+P_{41}u_4=0, \quad (101)$$

$$(\lambda_{31}-\lambda_4)(x_4x_{32}'+y_4y_{32}'+z_4z_{32}')+P_{42}u_4=0, \quad (102)$$

$$(\lambda_{33}-\lambda_4)(x_4x_{33}'+y_4y_{33}'+z_4z_{33}')+P_{43}u_4=0, \quad (103)$$

in which all the modes are rectified and are rectangular.

In virtue of the separation property the fourth-order equations will have at least one root equal to λ_{31} and for which the relative modes will be

$$x_{41}=x_{31}+px_{32}, \quad y_{41}=y_{31}+py_{32}, \quad z_{41}=z_{31}+pz_{32}, \quad u_{41}=0,$$

and which must satisfy the relation

$$(a_{14}-b_{14}\lambda_{31})x_{41}+(a_{24}-b_{24}\lambda_{31})y_{41}+(a_{34}-b_{34}\lambda_{31})z_{41}=0, \quad (104)$$

$$\text{or} \quad (P_{41})_{\lambda_{31}}+p(P_{42})_{\lambda_{31}}=0; \quad (105)$$

$$\text{so that} \quad p=-(P_{41}/P_{42})_{\lambda_{31}}. \quad (106)$$

The other three roots of the fourth-order equations are found from the escalator equation

$$(P_{41}^2+P_{42}^2)/(\lambda_{31}-\lambda_4)+P_{43}^2/(\lambda_{33}-\lambda_4)=u_{44}-b_{44}\lambda_4, \quad (107)$$

and their associated modes by the expressions

$$x_4/u_4=-[(P_{41}x_{31}+P_{42}x_{32})/(\lambda_{31}-\lambda_4)+P_{43}x_{33}/(\lambda_{33}-\lambda_4)], \quad (108)$$

$$x_4'/u_4=b_{14}-[(P_{41}x_{31}'+P_{42}x_{32}')/(\lambda_{31}-\lambda_4)+P_{43}x_{33}'/(\lambda_{33}-\lambda_4)], \quad (109)$$

and similarly for y_4/u_4 , z_4/u_4 , y_4'/u_4 , z_4'/u_4 .

If, however, $(P_{41})_{\lambda_{31}}=0$, $(P_{42})_{\lambda_{31}}=0$, then there is an additional root $\lambda_{42}=\lambda_{31}$ which satisfies the fourth-order equations. (106) will now be inapplicable, but it is clear by inspection that the following two sets of rectangular modes satisfy the equations, viz. :—

$$x_{41}=x_{31}, \quad y_{41}=y_{31}, \quad z_{41}=z_{31}=0, \quad u_{41}=0;$$

$$x_{42}=x_{32}, \quad y_{42}=y_{32}, \quad z_{42}=z_{32}, \quad u_{42}=0;$$

with their appropriate transposed modes. The two other roots and their associated modes are found in the usual way. If one of these roots is also equal to λ_{31} , let its rectified modes be x_{43} , y_{43} , z_{43} , u_{43} , x_{43}' , y_{43}' , z_{43}' , u_{43}' .

For the three equal roots λ_{41} of the fourth-order the equations will be satisfied by any linear combination of the rectangular modes found, *i. e.* any combination of the form $x_{41}+p_r y_{41}+q_r z_{41}$, etc., $x_{41}'+p_r y_{41}'+q_r z_{41}'$, etc., in which for rectification we divide by $\sqrt{1+p_r^2+q_r^2}$.

Taking r to be 1, 2, 3 respectively, the conditions for orthogonality will be

$$1+p_1p_2+q_1q_2=0; \quad 1+p_2p_3+q_2q_3=0; \quad 1+p_1p_3+q_1q_3=0. \quad (110)$$

It thus appears that if we have r equal roots for any particular order for which distinct sets of rectangular modes are known, then for the succeeding order there will be $r-1$ of these equal separated roots and for which mutually rectangular modes are directly ascertainable. If the P 's for the succeeding order, applicable to these equal roots, are zero, then there will be an additional root equal to the equal separated roots, and for which the modes can be found by inspection. Furthermore, one other such equal root may arise for which the modes may be found in the usual way.

11. Equal and/or Complex Roots for the Unsymmetrical Canonical Case.

In this case the theorem regarding separation of roots by successive orders does not necessarily hold good. Also, the latent roots are not necessarily real. To illustrate the escalator procedure in this case we shall consider the following fourth-order equations

$$(a_{11}-\lambda)x+a_{21}y+a_{31}z+a_{41}u=0, \quad . \quad . \quad . \quad (111)$$

$$a_{12}x+(a_{22}-\lambda)y+a_{32}z+a_{42}u=0, \quad . \quad . \quad . \quad (112)$$

$$a_{13}x+a_{23}y+(a_{33}-\lambda)z+a_{43}u=0, \quad . \quad . \quad . \quad (113)$$

$$a_{14}x+a_{24}y+a_{34}z+(a_{44}-\lambda)u=0. \quad . \quad . \quad . \quad (114)$$

We first take the second-order equations (111), (112), with $z=0$, $u=0$. Let the two latent roots be λ_{21} , λ_{22} and let their respective associated rectified modes be x_{21} , y_{21} , x_{21}' , y_{21}' ; x_{22} , y_{22} , x_{22}' , y_{22}' . For the third-order equations (111), (112), (113), with $u=0$ we have

$$P_{31}'=a_{31}x_{21}'+a_{32}y_{21}', \quad P_{32}'=a_{31}x_{22}'+a_{32}y_{22}'; \quad . \quad . \quad (115)$$

$$P_{31}=a_{13}x_{21}+a_{23}y_{21}, \quad P_{32}=a_{13}x_{22}+a_{23}y_{22}. \quad . \quad . \quad (116)$$

The escalator latent root equation for the third-order is given by

$$P_{31}P_{31}'/(\lambda_{21}-\lambda_3)+P_{32}P_{32}'/(\lambda_{22}-\lambda_3)=a_{33}-\lambda_3, \quad . \quad . \quad (117)$$

and the corresponding modes by

$$x_3/z_3 = -[P_{31}'x_{21}/(\lambda_{21}-\lambda_3)+P_{32}'x_{22}/(\lambda_{22}-\lambda_3)], \quad . \quad . \quad (118)$$

$$y_3/z_3 = -[P_{31}'y_{21}/(\lambda_{21}-\lambda_3)+P_{32}'y_{22}/(\lambda_{22}-\lambda_3)]; \quad . \quad . \quad (119)$$

$$x_3'/z_3' = -[P_{31}x_{21}'/(\lambda_{21}-\lambda_3)+P_{32}x_{22}'/(\lambda_{22}-\lambda_3)], \quad . \quad . \quad (120)$$

$$y_3'/z_3' = -[P_{31}y_{21}'/(\lambda_{21}-\lambda_3)+P_{32}y_{22}'/(\lambda_{22}-\lambda_3)]. \quad . \quad . \quad (121)$$

Now (117) may be written

$$\frac{(\lambda_{21}+\lambda_{22}-\lambda_3)(P_{31}P_{31}'+P_{32}P_{32}')-(\lambda_{21}P_{31}P_{31}'+\lambda_{22}P_{32}P_{32}')}{(\lambda_{21}-\lambda_3)(\lambda_{22}-\lambda_3)}=a_{33}-\lambda_3. \quad . \quad . \quad (122)$$

Also it may be shown that

$$P_{31}P_{31}'+P_{32}P_{32}'=a_{13}a_{31}+a_{23}a_{32}, \quad . \quad . \quad . \quad (123)$$

$$\begin{aligned} \lambda_{21}P_{31}P_{31}'+\lambda_{22}P_{32}P_{32}' &= a_{11}a_{13}a_{31}+a_{22}a_{23}a_{32} \\ &+ a_{12}a_{23}a_{31}+a_{21}a_{32}a_{13}. \quad . \quad . \quad (124) \end{aligned}$$

Hence, if $\lambda_{22}=\lambda_{21}$, (122) becomes

$$\begin{aligned} &[(2\lambda_{21}-\lambda_3)(a_{13}a_{31}+a_{23}a_{32})-(a_{11}a_{13}a_{31}+a_{22}a_{23}a_{32} \\ &+ a_{12}a_{23}a_{31}+a_{21}a_{32}a_{13})]/(\lambda_{21}-\lambda_3)^2=a_{33}-\lambda_3. \quad . \quad (125) \end{aligned}$$

and the modal expressions (118), (119), (120), (121) may be written

$$x_3/z_3 = -[(2\lambda_{21}-\lambda_3)a_{31}-(a_{11}a_{31}+a_{21}a_{32})]/(\lambda_{21}-\lambda_3)^2, \quad (126)$$

$$y_3/z_3 = -[(2\lambda_{21}-\lambda_3)a_{32}-(a_{12}a_{31}+a_{22}a_{32})]/(\lambda_{21}-\lambda_3)^2; \quad (127)$$

$$x_3'/z_3' = -[(2\lambda_{21}-\lambda_3)a_{13}-(a_{11}a_{13}+a_{12}a_{23})]/(\lambda_{21}-\lambda_3)^2, \quad (128)$$

$$y_3'/z_3' = -[(2\lambda_{21}-\lambda_3)a_{23}-(a_{21}a_{13}+a_{22}a_{23})]/(\lambda_{21}-\lambda_3)^2. \quad (129)$$

Now when the two roots of the second-order equations are equal the modes are in general indeterminate and may become infinite. But we have shown how we can proceed to the solution of the third-order equations without a detailed knowledge of these modes.

If the two roots of the second-order equations are complex, let them be $\lambda_{21} \pm i\mu_{21}$; then in place of $(\lambda_{21} - \lambda_3)^2$ we write $(\lambda_{21} - \lambda_3)^2 + \mu_{21}^2$ in (125), (126), (127), (128), (129).

We next consider the case where the two roots of the second-order equations are real and are either distinct or equal. Further, we will assume that the roots of the third-order equations are real and distinct. For the fourth-order equations we have, with the usual notation,

$$(\lambda_{31} - \lambda_4)(x_4 x_{31}' + y_4 y_{31}' + z_4 z_{31}') + P_{41}' u_4 = 0, \quad . \quad . \quad (130)$$

$$(\lambda_{32} - \lambda_4)(x_4 x_{32}' + y_4 y_{32}' + z_4 z_{32}') + P_{42}' u_4 = 0, \quad . \quad . \quad (131)$$

$$(\lambda_{33} - \lambda_4)(x_4 x_{33}' + y_4 y_{33}' + z_4 z_{33}') + P_{43}' u_4 = 0; \quad . \quad . \quad (132)$$

and for the transposed equations

$$(\lambda_{31} - \lambda_4)(x_4' x_{31} + y_4' y_{31} + z_4' z_{31}) + P_{41} u_4' = 0, \quad . \quad . \quad (133)$$

$$(\lambda_{32} - \lambda_4)(x_4' x_{32} + y_4' y_{32} + z_4' z_{32}) + P_{42} u_4' = 0, \quad . \quad . \quad (134)$$

$$(\lambda_{33} - \lambda_4)(x_4' x_{33} + y_4' y_{33} + z_4' z_{33}) + P_{43} u_4' = 0. \quad . \quad . \quad (135)$$

Thus, if λ_{31} is also a root of the fourth-order equations, then

$$P_{41}' = 0, u_{41}' = 0; \text{ or } P_{41} = 0, u_{41} = 0; \text{ or both.} \quad . \quad . \quad (136)$$

If $P_{41}' = 0$, we have associated with the root $\lambda_{41} = \lambda_{31}$ the relative transposed modes

$$x_{41}' = x_{31}', \quad y_{41}' = y_{31}', \quad z_{41}' = z_{31}', \quad u_{41}' = 0. \quad . \quad . \quad (137)$$

Suppose now that the other three roots of the fourth-order equations are distinct and different from λ_{41} , we may proceed to find the modes x_{41} , y_{41} , z_{41} , u_{41} as follows:—

For the modes when rectified

$$x_{41} x_{41}' + x_{42} x_{42}' + x_{43} x_{43}' + x_{44} x_{44}' = 1, \quad . \quad . \quad . \quad (138)$$

and similarly for the y , z modes.

Thus we may find x_{41} , y_{41} , z_{41} . Finally u_{41} may be found from (114) and will be given by

$$u_{41} = -(a_{14} x_{41} + a_{24} y_{41} + a_{34} z_{41}) / (a_{44} - \lambda_{41}). \quad . \quad . \quad (139)$$

We thus have the rectified values of all the modes pertaining to the root λ_{41} , which modes will be rectangular in relation to the other modes.

But it may happen that the fourth-order equations may have a second root λ_{41} , and for which the modes are ascertainable in the ordinary way. This, however, will not affect the suggested procedure.

If both P_{41}' , P_{41} are zero then λ_{31} is clearly a root of the fourth-order equations with rectified modes.

$$\begin{aligned} x_{41} &= x_{31}, & y_{41} &= y_{31}, & z_{41} &= z_{31}, & u_{41} &= 0; \\ x_{41}' &= x_{31}', & y_{41}' &= y_{31}', & z_{41}' &= z_{31}', & u_{41}' &= 0, \end{aligned} \quad . \quad . \quad (140)$$

12. Reduction of a Root to Zero.

The following artifice for the reduction of a root to zero whilst retaining the symmetry of the equations is due to Mr. T. S. Wilson. Consider the Lagrangian fourth-order equations (1), (2), (3), (4). If λ_{41} is a root then

$$x_{41}'x_4 + y_{41}'y_4 + z_{41}'z_4 + u_{41}'u_4 = 0, \quad . \quad . \quad . \quad (141)$$

where x_4, y_4, z_4, u_4 are modes other than $x_{41}, y_{41}, z_{41}, u_{41}$.

Now subtract $\lambda_{41}x_{41}'(x_{41}'x_4 + y_{41}'y_4 + z_{41}'z_4 + u_{41}'u_4)$
from equation (1); $\lambda_{41}y_{41}'(x_{41}'x_4 + y_{41}'y_4 + z_{41}'z_4 + u_{41}'u_4)$
from equation (2); and so on. We obtain

$$\begin{aligned} & [(a_{11} - \lambda_{41}x_{41}'^2) - b_{11}\lambda_4]x_4 + [(a_{12} - \lambda_{41}x_{41}'y_{41}') - b_{12}\lambda_4]y_4 \\ & + [(a_{13} - \lambda_{41}x_{41}'z_{41}') - b_{13}\lambda_4]z_4 + [(a_{14} - \lambda_{41}x_{41}'u_{41}') - b_{14}\lambda_4]u_4 = 0, \end{aligned} \quad (142)$$

and three similar equations.

If we put $x_4 = x_{41}, y_4 = y_{41}, z_4 = z_{41}, u_4 = u_{41}$ in these equations we have for the L.H.S. of (142) $-\lambda_4x_{41}'$. Similarly, for the other equations we obtain $-\lambda_4y_{41}', -\lambda_4z_{41}', -\lambda_4u_{41}'$. Hence these equations will have a zero latent root in place of the λ_{41} of the original equations, but with the original modes for this root. The other roots and their associated modes will be unaltered.

Clearly the reduction may be repeated for any number of known roots and their associated modes.

In connection with this artifice it may be proved that for the fourth-order equations (1), (2), (3), (4)

$$\lambda_{41}x_{41}'^2 + \lambda_{42}x_{42}'^2 + \lambda_{43}x_{43}'^2 + \lambda_{44}x_{44}'^2 = a_{11}, \quad . \quad . \quad . \quad (143)$$

$$\lambda_{41}y_{41}'^2 + \lambda_{42}y_{42}'^2 + \lambda_{43}y_{43}'^2 + \lambda_{44}y_{44}'^2 = a_{22}, \quad . \quad . \quad . \quad (144)$$

and so on;

$$\lambda_{41}x_{41}'y_{41}' + \lambda_{42}x_{42}'y_{42}' + \lambda_{43}x_{43}'y_{43}' + \lambda_{44}x_{44}'y_{44}' = a_{12}, \quad . \quad (145)$$

$$\lambda_{41}x_{41}'z_{41}' + \lambda_{42}x_{42}'z_{42}' + \lambda_{43}x_{43}'z_{43}' + \lambda_{44}x_{44}'z_{44}' = a_{13}, \quad . \quad (146)$$

$$\text{and so on; } x_{41}'^2 + x_{42}'^2 + x_{43}'^2 + x_{44}'^2 = b_{11}, \quad . \quad . \quad . \quad (147)$$

$$y_{41}'^2 + y_{42}'^2 + y_{43}'^2 + y_{44}'^2 = b_{22}, \quad . \quad . \quad . \quad (148)$$

$$\text{and so on; } x_{41}'y_{41}' + x_{42}'y_{42}' + x_{43}'y_{43}' + x_{44}'y_{44}' = b_{12}, \quad . \quad . \quad (149)$$

$$x_{41}'z_{41}' + x_{42}'z_{42}' + x_{43}'z_{43}' + x_{44}'z_{44}' = b_{13}, \quad . \quad . \quad (150)$$

and so on; in which the modes involved are the rectified values.

For the corresponding unsymmetrical canonical case we shall have

$$\lambda_{41}x_{41}x_{41}' + \lambda_{42}x_{42}x_{42}' + \lambda_{43}x_{43}x_{43}' + \lambda_{44}x_{44}x_{44}' = a_{11}, \quad . \quad . \quad (151)$$

$$\lambda_{41}x_{41}y_{41}' + \lambda_{42}x_{42}y_{42}' + \lambda_{43}x_{43}y_{43}' + \lambda_{44}x_{44}y_{44}' = a_{21}, \quad . \quad . \quad (152)$$

$$\lambda_{41}x_{41}'y_{41} + \lambda_{42}x_{42}'y_{42} + \lambda_{43}x_{43}'y_{43} + \lambda_{44}x_{44}'y_{44} = a_{12}, \quad . \quad . \quad (153)$$

and so on.

13. *Latent Root Equations arising from Solution of Linear Simultaneous Equations by Iteration Processes.*

Consider the following linear simultaneous equations of the third-order, viz.

$$a_{11}x + a_{21}y + a_{31}z = d_1, \quad . \quad . \quad . \quad . \quad . \quad (154)$$

$$a_{12}x + a_{22}y + a_{32}z = d_2, \quad . \quad . \quad . \quad . \quad . \quad (155)$$

$$a_{13}x + a_{23}y + a_{33}z = d_3; \quad . \quad . \quad . \quad . \quad . \quad (156)$$

and suppose the variables are iterated on the scheme

$$a_{11}x_{n+1} + a_{21}y_n + a_{31}z_n = d_1, \quad . \quad . \quad . \quad . \quad . \quad (157)$$

$$a_{12}x_{n+1} + a_{22}y_{n+1} + a_{32}z_n = d_2, \quad . \quad . \quad . \quad . \quad . \quad (158)$$

$$a_{13}x_{n+1} + a_{23}y_{n+1} + a_{33}z_{n+1} = d_3, \quad . \quad . \quad . \quad . \quad . \quad (159)$$

in which we choose any values y_0, z_0 in (157) and ascertain the corresponding x_1 ; then using z_0 and x_1 in (158) we obtain y_1 ; then using x_1 and y_1 in (159) we obtain z_1 ; and so on.

$$\text{Let} \quad \xi_n = x_{n+1} - x_n, \quad \eta_n = y_{n+1} - y_n, \quad \zeta_n = z_{n+1} - z_n, \quad . \quad . \quad . \quad (160)$$

$$\text{then we have} \quad a_{11}\xi_{n+1} + \xi_{21}\eta_n + a_{31}\zeta_n = 0, \quad . \quad . \quad . \quad . \quad . \quad (161)$$

$$a_{12}\xi_{n+1} + a_{22}\eta_{n+1} + a_{32}\zeta_n = 0, \quad . \quad . \quad . \quad . \quad . \quad (162)$$

$$a_{13}\xi_{n+1} + a_{23}\eta_{n+1} + a_{33}\zeta_{n+1} = 0. \quad . \quad . \quad . \quad . \quad . \quad (163)$$

For a solution of the finite difference equations (161), (162), (163), we assume

$$\xi_n = x\lambda^n, \quad \eta_n = y\lambda^n, \quad \zeta_n = z\lambda^n; \quad . \quad . \quad . \quad . \quad . \quad (164)$$

where x, y, z are arbitrary constants in relation to their associated λ . Substituting the values (164) in the equations (161), (162), (163) we obtain the latent root equations

$$a_{11}\lambda x + a_{22}y + a_{31}z = 0, \quad . \quad . \quad . \quad . \quad . \quad (165)$$

$$a_{12}\lambda x + a_{22}\lambda y + a_{32}z = 0, \quad . \quad . \quad . \quad . \quad . \quad (166)$$

$$a_{13}x + a_{23}y + a_{33}z = 0. \quad . \quad . \quad . \quad . \quad . \quad (167)$$

These equations will have two roots in λ . Let them be $\lambda_{21}, \lambda_{22}$ and let $x_{21}, y_{21}, z_{21}; x_{22}, y_{22}, z_{22}$ be their associated relative modes, of which for convenience we shall regard the z 's as unity. The solutions of equations (161), (162), (163) will therefore be given by

$$\xi_n = A_1 x_{21} \lambda_{21}^n + A_2 x_{22} \lambda_{22}^n, \quad . \quad . \quad . \quad . \quad . \quad (168)$$

$$\eta_n = A_1 y_{21} \lambda_{21}^n + A_2 y_{22} \lambda_{22}^n, \quad . \quad . \quad . \quad . \quad . \quad (169)$$

$$\zeta_n = A_1 \lambda_{21}^n + A_2 \lambda_{22}^n, \quad . \quad . \quad . \quad . \quad . \quad (170)$$

where A_1, A_2 are arbitrary constants whose values depend on the initial values assumed in the iteration.

Clearly ξ_n, η_n, ζ_n will be convergent provided the moduli of $\lambda_{21}, \lambda_{22}$ are less than unity.

If the coefficients of the simultaneous equations are such that $a_{sr} = a_{rs}$, and if further they are coefficients of a positive homogeneous quadratic

function, then these conditions are fulfilled. Various proofs have been given of this, but the following, which is due to Prof. G. Temple, appears to be the most concise³—

In the equations (165), (166), (167), let $a_{sr}=a_{rs}$,

$$\text{and let} \quad x=p_1+iq_1, \quad y=p_2+iq_2, \quad z=p_3+iq_3; \quad \dots \quad (171)$$

$$\text{also let} \quad \bar{x}=p_1-iq_1, \quad \bar{y}=p_2-iq_2, \quad \bar{z}=p_3-iq_3. \quad \dots \quad (172)$$

Multiply (165), (166) and $\lambda \times$ (167) by \bar{x} , \bar{y} , \bar{z} respectively and add. We obtain

$$\begin{aligned} & \lambda[\Sigma a_{rr}(p_r^2+q_r^2)+\Sigma a_{rs}(p_r p_s+q_r q_s)+i\Sigma a_{rs}(p_r q_s-p_s q_r)] \\ & +\Sigma a_{rs}(p_r p_s+q_r q_s)-i\Sigma a_{rs}(p_r q_s-p_s q_r)=0. \quad \dots \quad (173) \end{aligned}$$

$$\text{Hence} \quad [\text{mod. } \lambda]^2=(P^2+Q^2)/[(P+R)^2+Q^2], \quad \dots \quad (174)$$

$$\text{where} \quad P=\Sigma a_{rs}(p_r p_s+q_r q_s), \quad \dots \quad (175)$$

$$Q=\Sigma a_{rs}(p_r q_s-p_s q_r), \quad \dots \quad (176)$$

$$R=\Sigma a_{rr}(p_r^2+q_r^2). \quad \dots \quad (177)$$

$$\text{Now} \quad (P+R)^2=P^2+R(R+2P) \quad \dots \quad (178)$$

$$\text{and} \quad R+2P=\Sigma a_{rr}(p_r^2+q_r^2)+2\Sigma a_{rs}(p_r p_s+q_r q_s). \quad \dots \quad (179)$$

Hence, if the coefficients a_{rs} are coefficients of a positive homogeneous quadratic function, $R+2P$ as well as R will be positive, and in these circumstances the modulus of λ will be less than unity. It also follows that if λ is real its numerical value must be less than unity.

It may be noted that any set of equations like (154), (155), (156) may be transformed into a homogeneous "normalized" set as follows:— Multiply (154), (155), (156) respectively by a_{11} , a_{12} , a_{13} and add; then likewise by a_{21} , a_{22} , a_{23} and a_{31} , a_{32} , a_{33} . Thus, since

$$(a_{11}x+a_{21}y+a_{31}z)^2+(a_{12}x+a_{22}y+a_{32}z)^2+(a_{13}x+a_{23}y+a_{33}z)^2$$

is a positive homogeneous quadratic function, it follows that the coefficients, resulting from normalizing in the manner described, will be coefficients of such a function.

It is due to this theorem that the iteration process described was found to be convergent when applied to "normal" equations; and more recently that "moment distribution"⁽³⁾ and "relaxation"⁽⁴⁾ methods, which are really physical applications of algebraic iteration, proved practical propositions.

The convergency of an iteration process depends *inter alia* on the scheme of iteration. For example, if we iterate the equations (154), (155), (156), supposed homogeneous, on the scheme

$$a_{11}x_{n+1}+a_{12}y_n+a_{13}z_n=d_1, \quad \dots \quad (180)$$

$$a_{12}x_n+a_{22}y_{n+1}+a_{23}z_n=d_2 \quad \dots \quad (181)$$

$$a_{13}x_n+a_{23}y_n+a_{33}z_{n+1}=d_3, \quad \dots \quad (182)$$

then the appropriate latent root equations will be given by

$$a_{11}\lambda x+a_{12}y+a_{13}z=0, \quad \dots \quad (183)$$

$$a_{12}x + a_{22}\lambda y + a_{23}z = 0, \quad . \quad . \quad . \quad . \quad . \quad (184)$$

$$a_{13}x + a_{23}y + a_{33}\lambda z = 0, \quad . \quad . \quad . \quad . \quad . \quad (185)$$

for which the roots in λ will be real provided a_{rr} is real and positive and a_{rs} is real. Multiply these equations by x, y, z respectively and add. We obtain

$$\lambda R + 2P = 0, \quad . \quad . \quad . \quad . \quad . \quad (186)$$

where

$$P = \Sigma a_{12}xy, \quad R = \Sigma a_{11}x^2. \quad . \quad . \quad . \quad . \quad (187)$$

Thus, if the coefficients a_{rs} are coefficients of a positive homogeneous quadratic function, $R + 2P$ will be positive; and if P is negative, $R > 2P$; hence if λ is positive it must be less than unity; but if λ is negative it can have any numerical value whatever.

Thus the iteration scheme (180), (181), (182) will not necessarily be convergent when the a 's are coefficients of a positive homogeneous quadratic function.

In the foregoing we have dealt with simultaneous equations of only three variables, but it is clear that the propositions proved are quite general whatever the number of variables involved.

14. *The Solution of Canonical Lagrangian Equations by means of Linear Simultaneous Equations.*

It will now be shown that it is possible to arrive at a solution of canonical Lagrangian frequency equations by means of linear simultaneous equations, but it is doubtful whether the method is as efficient as the escalator process for actual computation.

To illustrate the method we consider the following third-order canonical equations for which a solution is required:

$$(a_{11} - \lambda)x + a_{21}y + a_{31}z = 0, \quad . \quad . \quad . \quad . \quad (188)$$

$$a_{12}x + (a_{22} - \lambda)y + a_{32}z = 0, \quad . \quad . \quad . \quad . \quad (189)$$

$$a_{13}x + a_{23}y + (a_{33} - \lambda)z = 0. \quad . \quad . \quad . \quad . \quad (190)$$

Let us border these equations as follows:

$$(a_{11} - \lambda)x + a_{21}y + a_{31}z + a_{41}u = 0, \quad . \quad . \quad . \quad . \quad (191)$$

$$a_{12}x + (a_{22} - \lambda)y + a_{32}z + a_{42}u = 0, \quad . \quad . \quad . \quad . \quad (192)$$

$$a_{13}x + a_{23}y + (a_{33} - \lambda)z + a_{43}u = 0, \quad . \quad . \quad . \quad . \quad (193)$$

$$0 + 0 + z + (a_{44} - \lambda)u = 0, \quad . \quad . \quad . \quad . \quad (194)$$

where $a_{41}, a_{42}, a_{43}, a_{44}$ are as yet arbitrary,

but

$$a_{14} = 0, \quad a_{24} = 0, \quad a_{34} = 1.$$

Choosing any value λ_r , we can solve (191), (192), (193) as linear simultaneous equations and obtain a solution of the form

$$x_r/u_r = (\alpha_{11})_r a_{41} + (\alpha_{21})_r a_{42} + (\alpha_{31})_r a_{43}, \quad . \quad . \quad . \quad (195)$$

$$y_r/u_r = (\alpha_{12})_r a_{41} + (\alpha_{22})_r a_{42} + (\alpha_{32})_r a_{43}, \quad . \quad . \quad . \quad (196)$$

$$z_r/u_r = (\alpha_{13})_r a_{41} + (\alpha_{23})_r a_{42} + (\alpha_{33})_r a_{43}. \quad . \quad . \quad . \quad (197)$$

Since λ_r is at our choice we can so choose it as to make the diagonal coefficients like $(a_{11}-\lambda_r)$ as heavily loaded as we please in relation to the other coefficients, and thereby arrange for the rapid convergency of the ordinary iteration process when applied to the solution of the linear simultaneous equations involved.

Substituting (197) in (194) we obtain

$$(\alpha_{13})_r a_{41} + (\alpha_{23})_r a_{42} + (\alpha_{33})_r a_{43} + (a_{44} - \lambda_r) = 0. \quad (198)$$

Having chosen four values of λ_r , we have four linear equations of the form (198) for the determination of a_{41} , a_{42} , a_{43} , a_{44} . We notice, however, that

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = a_{11} + a_{22} + a_{33} + a_{44}; \quad (199)$$

hence a_{44} is immediately known.

We thus require the solution of any three of the four linear simultaneous equations (198) ($r=1, 2, 3, 4$) to find a_{41} , a_{42} , a_{43} . (A difficulty here is that we do not know whether the coefficients $(\alpha_{13})_r$, etc. will be such as to be consistent with convergence of the ordinary iteration process.)

Next we consider the transposed equations

$$(a_{11} - \lambda)x' + a_{12}y' + a_{13}z' = 0, \quad (200)$$

$$a_{21}x' + (a_{22} - \lambda)y' + a_{23}z' = 0, \quad (201)$$

$$a_{31}x' + a_{32}y' + (a_{33} - \lambda)z' + u' = 0, \quad (202)$$

$$a_{41}x' + a_{42}y' + a_{43}z' + (a_{44} - \lambda)u' = 0. \quad (203)$$

For the value λ_r , equations (200), (201), (202), all have solutions of the form

$$x'_r/u'_r = (\alpha_{13})_r, \quad y'_r/u'_r = (\alpha_{23})_r, \quad z'_r/u'_r = (\alpha_{33})_r. \quad (204)$$

For the rectification of the modes we must have

$$x_r x'_r + y_r y'_r + z_r z'_r + u_r u'_r = 1. \quad (205)$$

Hence the rectified value of $u_r u'_r$ is given by

$$\begin{aligned} \frac{1}{u_r u'_r} = & 1 + [(\alpha_{13})_r (\alpha_{11})_r + (\alpha_{23})_r (\alpha_{12})_r + (\alpha_{33})_r (\alpha_{13})_r] a_{41} \\ & + [(\alpha_{13})_r (\alpha_{21})_r + (\alpha_{23})_r (\alpha_{22})_r + (\alpha_{33})_r (\alpha_{23})_r] a_{42} \\ & + [(\alpha_{13})_r (\alpha_{31})_r + (\alpha_{23})_r (\alpha_{32})_r + (\alpha_{33})_r (\alpha_{33})_r] a_{43}. \end{aligned} \quad (206)$$

Next, making use of the escalator in reverse for the equations (191), (192), (193), (194) we have for the λ equation pertaining to the third-order equations (188), (189), (190).

$$\sum_{r=1}^{r=4} u_r u'_r / (\lambda_r - \lambda) = 0. \quad (207)$$

For the modes of the third-order equations associated with a root of (207) we have

$$\begin{aligned} x/P &= \Sigma x_r u'_r / (\lambda_r - \lambda), & y/P &= \Sigma y_r u'_r / (\lambda_r - \lambda), \\ z/P &= \Sigma z_r u'_r / (\lambda_r - \lambda), & x'/P' &= \Sigma x'_r u_r / (\lambda_r - \lambda), \\ y'/P' &= \Sigma y'_r u_r / (\lambda_r - \lambda), & z'/P' &= \Sigma z'_r u_r / (\lambda_r - \lambda), \end{aligned} \quad (208)$$

where $P=z, P'=a_{41}x'+a_{42}y'+a_{43}z', \dots \dots \dots (209)$

and to rectify which, we have

$$1/P^2 = \pm \Sigma u_r u_r' / (\lambda_r - \lambda)^2, \dots \dots \dots (210)$$

in which PP' is taken to $\pm P^2$, the sign being chosen so as to make P^2 positive.

If the equations (188), (189), (190) are symmetrical, then there is a simplification in that

$$(\alpha'_{pq})_r = (\alpha_{pq})_r \quad \text{and} \quad (\alpha_{pq})_r = (\alpha_{qp})_r. \dots \dots \dots (211)$$

References.

- (1) W. J. Duncan and A. R. Collar, "On a Method for the Solution of Oscillation Problems by Matrices," Phil. Mag. vol. xvii. p. 865 (1934).
- (2) J. Morris and J. W. Head, "Lagrangian Frequency Equations; an 'Escalator' Method for Numerical Solution," Aircraft Engineering, November, 1942.
- (3) H. Cross, "Analysis of Continuous Frames by distributing Fixed End Moments," Proc. Amer. Soc. C. E., May, 1930.
- (4) R. V. Southwell, "Relaxation Methods in Engineering Science." Clarendon Press, 1940.
- (5) J. Morris, "A Successive Approximation Process for solving Simultaneous Linear Equations," Aer. Res. Cttee., Rep. and Mem. No. 1711 (1936).

APPENDIX.

Numerical Example of Application of Escalator Process.

Suppose we wish to solve the following sixth-order symmetrical canonical equations, viz.:—

<i>a.</i>	<i>b.</i>	<i>c.</i>	<i>d.</i>	<i>e.</i>	<i>f.</i>
5401.79— λ	5232.86	4357.85	3622.42	2764.72	1846.91 =0
5232.86	7873.70— λ	3622.42	5256.51	1846.91	2802.69 =0
4357.85	3622.42	3883.21— λ	2973.04	2639.74	1679.36 =0
3622.42	5256.51	2973.04	4378.57— λ	1679.36	2632.46 =0
2764.72	1846.91	2639.74	1679.36	2017.58— λ	1149.21 =0
1846.91	2802.69	1679.36	2632.46	1149.21	1942.45— λ =0

We start by solving the quadratic in λ for the second-order equations, *i. e.*,

<i>a.</i>	<i>b.</i>
5401.79— λ	5232.86 =0
5232.86	7873.70— λ =0

for which we find the two roots

$$\lambda_{21}=12014\cdot585, \quad \lambda_{22}=1260\cdot905.$$

For $\lambda_{21}=12014\cdot585$, we find that the relative modes are $a_{21}=1$, $b_{21}=1\cdot263706$, and to rectify these we have to divide by the "divisor"

$$\delta_{21}=\sqrt{a_{21}^2+b_{21}^2}=1\cdot611506.$$

For $\lambda_{22}=1260\cdot905$, we find that the relative modes are $a_{22}=-1\cdot263706$, $b_{22}=1$, and to rectify we divide by $\delta_{22}=\delta_{21}=1\cdot611506$.

We next proceed to solve the third-order equations:—

<i>a.</i>	<i>b.</i>	<i>c.</i>
5401·79— λ	5232·86	4357·85 =0
5232·86	7873·70— λ	3622·42 =0
4357·85	3622·42	3883·21— λ =0

First we evaluate

$$P_{31}=\frac{4357\cdot85 \times 1 + 3622\cdot42 \times 1\cdot263706}{1\cdot611506} = 5544\cdot828,$$

$$P_{32}=\frac{4357\cdot85 \times (-1\cdot263706) + 3622\cdot42 \times 1}{1\cdot611506} = -1169\cdot478.$$

We note that

$$P_{31}a_{21}/\delta_{21} + P_{32}a_{22}/\delta_{22} = 4357\cdot852,$$

$$P_{31}b_{21}/\delta_{21} + P_{32}b_{22}/\delta_{22} = 3622\cdot422.$$

Having found the P's and carried out these checks, we form the escalator latent root equation for the third-order, viz:

$$P_{31}^2/(12014\cdot585 - \lambda) + P_{32}^2/(1260\cdot905 - \lambda) = 3883\cdot21 - \lambda,$$

and proceed to find the roots in λ as follows:—

λ .	15000.	14838·5.	14845·15.	14845·274.	14845·272.
12014·585— λ	— 2985·415	— 2823·91	— 2830·56	— 2830·684	— 2830·682
1260·905— λ	— 13739·095	— 13577·59	— 13584·24	— 13584·364	— 13584·362
3883·21— λ	— 11116·79	— 10955·29	— 10961·94	— 10962·064	— 10962·062
$P_{31}^2/(12014\cdot585 - \lambda)$.	— 10298·43	— 10887·44	— 10861·86	— 10861·374	— 10861·382
$P_{32}^2/(1260\cdot905 - \lambda)$.	— 99·55	— 100·73	— 100·68	— 100·681	— 100·681
A	— 718·81	+ 32·88	+ 0·60	— 0·009	+ 0·001
$P_{31}^2/(12014\cdot585 - \lambda)^2$	3·45	3·86	3·837	3·837	3·837020
$P_{32}^2/(1260\cdot905 - \lambda)^2$	0·007	0·01	0·007	0·007	0·007412
D ²	4·45	4·87	4·844	4·844	4·844432
$\Delta\lambda = A/D^2$	— 161·5	+ 6·65	+ 0·124	— 0·002	+ 0·0002

where

$$A = (3883 \cdot 21 - \lambda) - P_{31}^2 / (12014 \cdot 585 - \lambda) - P_{32}^2 / (1260 \cdot 905 - \lambda),$$

$$\text{and} \quad D^2 = 1 + P_{31}^2 / (12014 \cdot 585 - \lambda)^2 + P_{32}^2 / (1260 \cdot 905 - \lambda)^2.$$

In the table λ is an approximation steadily approaching the true value 14845.272. The first "shot" 15000 is a pure guess, but from the form of the escalator equation it is easy to make this first guess close enough for the Newtonian successive approximation process here used to reach a sufficiently accurate result after not more than four rounds. We know that the highest root sought must be greater than 12014.585, while the lowest root is below 1260.905. We know also that if in any one of the ranges $\lambda > 12014 \cdot 585$; $12014 \cdot 585 > \lambda > 1260 \cdot 905$; and $\lambda < 1260 \cdot 905$, we increase λ the corresponding value of the R.H.S. of the escalator equation is decreased while that of the L.H.S. is increased. If, therefore, the first guess is very much out it is easy to see in which direction we may make a better guess.

The second approximation to the root is $15000 + A/D^2$ and the rest of the table is self-explanatory. Great accuracy in the calculation of A and D^2 is not required until the approximation is very close to the actual root; but when the root is found D^2 must be calculated to a sufficiently high degree of accuracy for comparison with the "divisor" δ for the rectification of the modes, which in the particular case will be given by

$$a_3/c_3 = -[P_{31}a_{21}/(12014 \cdot 585 - \lambda)\delta_{21} + P_{32}a_{22}/(1260 \cdot 905 - \lambda)\delta_{22}],$$

$$b_3/c_3 = -[P_{31}b_{21}/(12014 \cdot 585 - \lambda)\delta_{21} + P_{32}b_{22}/(1260 \cdot 905 - \lambda)\delta_{22}],$$

in which we regard c_3 as unity.

Having obtained satisfactory agreement between D_r^2 and δ_r^2 , we use D_r for the divisor in preference to δ_r .

After all the roots of the escalator equation have been found, and before we ascertain their associated modes, a useful check may be applied. This is based on the fact that for canonical equations the sum of the λ roots for any particular order is equal to the sum of the elements of the leading diagonal.

If the three checks indicated are found to be satisfactory, then it is considered unnecessary to go through the laborious process of checking each set of values in the equations.

Following the procedure described the results for the third, fourth, fifth and sixth-orders were found to be—

Third Order.

r .	λ_{3r} .	a_{3r} .	b_{3r} .	c_{3r} .	δ_{3r}^2 .	D_{3r}^2 .
1	14845.272	1.283038	1.482647	1	4.844285	4.844432
2	2202.440	0.623359	-1.213905	1	2.8621418	2.862142
3	110.992	-1.086572	0.265818	1	2.251298	2.251298

Fourth Order.

$r.$	$\lambda_{4r}.$	$a_{4r}.$	$b_{4r}.$	$c_{4r}.$	$d_{4r}.$	$\delta_{4r}^2.$	$D_{4r}^2.$
1	18361.247	1.134132	1.374739	0.8900681	1	4.969475	4.969487
2	2436.398	-1.489491	1.726257	-1.890544	1	9.772703	9.772692
3	727.339	-0.459484	-0.581534	0.359946	1	1.678868	1.678889
4	12.2919	1.962798	-1.060243	-1.985574	1	9.919195	9.919243

Fifth Order.

$r.$	$\lambda_{5r}.$	$a_{5r}.$	$b_{5r}.$	$c_{5r}.$	$d_{5r}.$	$e_{5r}.$	$\delta_{5r}^2.$	$D_{5r}^2.$
1	19455.13	2.057657	2.405581	1.642591	2.768411	1	16.846155	16.846193
2	3199.41	0.663669	-1.195691	0.991125	-0.631812	1	4.251649	4.251668
3	778.438	-2.067195	-1.620615	0.626662	3.462032	1	20.278059	20.278381
4	115.331	-0.452164	0.43411	-0.366802	-0.289181	1	1.611073	1.611074
5	6.5463	2.946028	-1.442696	-3.779099	1.479357	1	28.230539	28.230712

Sixth Order.

$r.$	$\lambda_{6r}.$	$a_{6r}.$	$b_{6r}.$	$c_{6r}.$	$d_{6r}.$	$e_{6r}.$	$f_{6r}.$
1	20617.56	1.981943	2.352583	1.591799	1.757966	0.974624	1
2	3288.483	-1.983156	2.826956	-2.645322	1.696344	-2.555482	1
3	1366.676	-0.585281	-0.672433	0.139023	0.649407	0.388791	1
4	193.974	-0.050247	0.671437	-0.689469	-1.339519	0.99773	1
5	28.447	1.276825	-0.609134	-0.322565	-0.277764	-1.124037	1
6	2.1598	-1.556846	0.816074	3.056788	-1.686259	-1.781166	1

Order of equation.	Sum of latent roots found by escalation.	Sum of coefficients in leading diagonal.
3rd.	17158.704	17158.70
4th.	21537.2759	21537.27
5th.	23554.8553	23554.85
6th.	25497.2998	25497.30

LXXXVII. *Thermal Transpiration and Reversible Process for a Degenerate Gas.*

By D. V. GOGATE, M.Sc., Ph.D. (London) and V. N. UPADHYAYA, B.Sc.,
Physics Department, Baroda College, Baroda (India)*.

[Received August 27, 1943.]

Introduction.

SOME time ago Kennard showed in an important paper⁽¹⁾ that the basic theorem of thermodynamics that "The entropy-change of the Universe (*i. e.* the system and its surroundings) during a reversible process is zero" does not possess that universal validity which is characteristic of theorems in thermodynamics. He has given two examples of reversible processes where the above theorem is not applicable. One of these deals with the phenomena of thermal transpiration for a perfect classical gas and the other with a thermo-electric couple. In the case of thermal transpiration there takes place all the time an irreversible flow of heat from the hotter to the colder chamber, and this flow cannot be stopped without stopping thermal transpiration itself. Any reversible process that we shall consider between two chambers in equilibrium under thermal transpiration will be, unavoidably and inevitably, accompanied by an irreversible process. The change of entropy of the Universe, in the reversible process, will not be zero, but this could not be interpreted as a violation of the second law for, the reversible process, even in principle, could not be isolated from the irreversible process which accompanies it, and the entropy-increase due to the irreversible process would more than compensate any decrease of entropy due to the reversible process.

"An isolable or independent reversible process never changes the entropy of the Universe, but this is not true for a reversible process which cannot, even in principle, be separated from an irreversible process." This will be referred to hereafter as Kennard's theorem.

The theorem of entropy-change during a reversible process is so important that it is of some interest to consider other possible examples where the usual theorem, *i. e.* without the modification introduced by Kennard, becomes invalid.

Kennard considered thermal transpiration for a classical perfect gas. In view of the recent studies of effusion and thermal transpiration phenomena for a degenerate gas, the discussion of Kennard can be extended to include the case of degeneracy. This is done in the present paper.

* Communicated by the Authors.

Consider two chambers A and B at temperatures T_1 and T_2 respectively ($T_1 > T_2$) containing a degenerate gas, degenerate in the sense of Fermi-Dirac statistics. The chambers communicate through a small orifice O and equilibrium conditions pertaining to thermal transpiration exist (fig. 1). The conditions which must be fulfilled will be

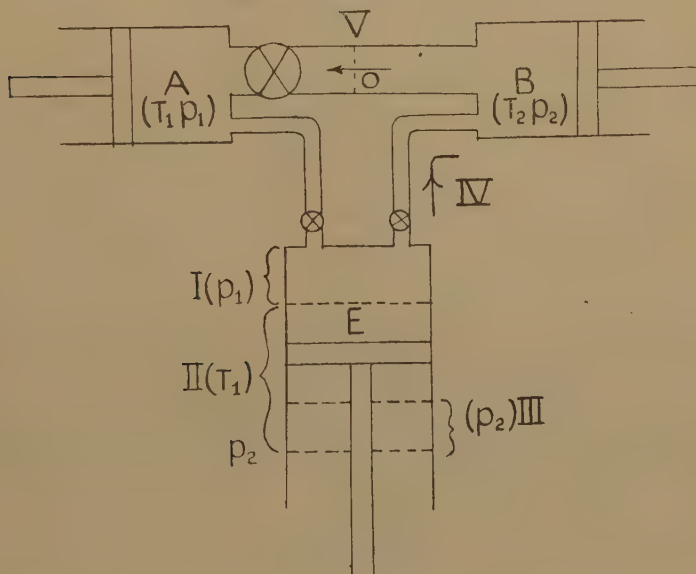
$$\frac{n_1^{4/3}}{n_2^{4/3}} = \frac{1 + \frac{\pi^2 k^2}{6\alpha^2} \frac{T_2^2}{n_2^{4/3}}}{1 + \frac{\pi^2 k^2}{6\alpha^2} \frac{T_1^2}{n_1^{4/3}}}, \quad \dots \quad (1)$$

or $6\alpha^2 n_1^{4/3} + \pi^2 k^2 T_1^2 = 6\alpha^2 n_2^{4/3} + \pi^2 k^2 T_2^2 = \text{constant}, \quad \dots \quad (2)$

where $\alpha = \frac{h^2}{2m} \left(\frac{3}{4\pi g} \right)^{2/3}, \quad \dots \quad (3)$

$n = N/V$ = number of particles per unit volume and h is Planck's constant.

Fig. 1.



The chambers are also connected through an expansion engine E, which allows the gas to be brought from B to A in a reversible way. Now let us circulate a gram of gas, reversibly in the usual slow way without appreciably disturbing the equilibrium. Let us remove the gas from the chamber A at temperature T_1 , lower its temperature to T_2 and its pressure to p_2 , and then restore it to the chamber B at temperature T_2 , after which it will flow automatically back through the small orifice O into the chamber A. We shall calculate the total change in entropy arising from this "reversible process."

We know that the energy of a degenerate gas in the non-relativistic case is given by ⁽²⁾

$$E_{\text{deg.}}^{\text{non-rel.}} = \frac{3}{5} \frac{N^{5/3} \alpha}{V^{2/3}} \left[1 + \frac{5}{12} \frac{\pi^2 k^2 T^2 V^{4/3}}{\alpha^2 N^{4/3}} \right] \quad (12)$$

Hence

$$\frac{5}{3} \frac{dE}{T} = -\frac{2}{3} \alpha N^{5/3} V^{-5/3} \frac{dV}{T} + \frac{5}{18} \frac{N^{1/3}}{\alpha} \pi^2 k^2 T V^{-1/3} dV + \frac{5}{6\alpha} N^{1/3} \pi^2 k^2 V^{2/3} dT \quad (13)$$

Differentiating (2), treating N_1 , V_1 and T_1 as constants, we get

$$\pi^2 k^2 T dT = \frac{4\alpha^2 N^{4/3} dV}{V^{7/3}}, \quad (14)$$

$$\text{whence} \quad dV = \frac{\pi^2 k^2 V^{7/3} T dT}{4\alpha^2 N^{4/3}} \quad (15)$$

Again from (2) we have

$$V = \frac{(6\alpha^2) N^{3/4}}{(\beta - \pi^2 k^2 T^2)^{3/4}}, \quad (16)$$

$$\text{where} \quad \beta = \left(\frac{6\alpha^2 N_1^{4/3}}{V_1^{4/3}} + \pi^2 k^2 T_1^2 \right) \quad (17)$$

Substituting (15) and (16) in equation (13) we get

$$\frac{5}{3} \frac{dE}{T} = -\frac{1}{6^{1/2}} \frac{N \pi^2 k^2 dT}{(\beta - \pi^2 k^2 T^2)^{1/2}} + \frac{5}{6^{1/2}} \frac{N \pi^2 k^2 dT}{(\beta - \pi^2 k^2 T^2)^{1/2}} + \frac{5 \times 6^{3/2}}{72} \frac{\pi^4 k^4 T^2 dT N}{(\beta - \pi^2 k^2 T^2)^{3/2}}, \quad (18)$$

and as $\pi^2 k^2 T^2$ is too small compared to β ,

$$\frac{5}{3} \frac{dE}{T} = \frac{4N \pi^2 k^2 dT}{(6\beta)^{1/2}} + \frac{5 \times 6^{3/2}}{72} \frac{\pi^4 k^4 T^2 N dT}{\beta^{3/2}}$$

Substituting the approximate value of $\beta \left(= \frac{6\alpha^2 N_1^{4/3}}{V_1^{4/3}} \right)$ and integrating, we get

$$\frac{5}{3} \int_{T_1}^{T_2} \frac{dE}{T} = \frac{2}{3\alpha} N^{1/3} \pi^2 k^2 V_1^{2/3} (T_1 - T_2) + \frac{5}{216} \frac{\pi^4 k^4 V_1^2}{\alpha^3 N} (T_1^3 - T_2^3) \quad (19)$$

or

$$S_5 = \frac{5}{3} \int_{T_1}^{T_2} \frac{dE}{T} = \frac{2}{3} \frac{N^{1/3} \pi^2 k^2 V_1^{2/3}}{\alpha} (T_1 - T_2) \text{ approx.} \quad (19)$$

In the case of a non-degenerate gas,

$$S_5 = (R + C_v) \log \frac{T_1}{T_2} \quad (20)$$

The total loss of entropy by the surroundings due to this cyclic transfer of the gas is thus

$$S = \Sigma S_n = S_2 + S_3 + S_5 = \frac{N^{1/3} \pi^2 k^2 V^{2/3}}{6\alpha} (T_1 - T_2), \quad (21)$$

since * $V_3=V_1=V$ to a first approximation, or

$$S = \frac{1}{3} \frac{R\pi^2 km}{h^2} \left(\frac{4\pi g}{3n} \right)^{2/3} (T_1 - T_2). \quad (22)$$

In the non-degenerate case,

$$S = S_2 = R \log p_1/p_2. \quad (23)$$

The gas itself, however, has the same entropy at the end as at the beginning. The entropy of the Universe has thus been decreased by a reversible process.

We have, however, not taken into consideration, in the above reasoning, the increase of entropy due to the conduction of heat from one chamber to the other through the gas in the plug, and it can be easily shown that the irreversible gain in entropy due to this cause is always ample to offset any irreversible decrease caused by circulating the gas with the pump.

We thus have before us an example of a reversible process which makes invalid the general theorem, viz. :—"A reversible process never changes the entropy of the Universe" if the restriction imposed by Kennard is ignored.

It may be of interest, further, to note that the ratio of the entropy-change during the reversible process to the difference between the entropies of the gas at temperatures T_1 and T_2 respectively, is given by

$$\frac{\Delta S}{s_1 - s_2} = \frac{1}{3} \text{ for a degenerate gas,}$$

and
$$\frac{\Delta S}{s_1 - s_2} = \frac{1}{4} \text{ for a non-degenerate gas,}$$

where Δs is the entropy-change during the reversible process discussed above. This ratio differs from unity in the case of degeneracy as well as non-degeneracy, but for a strictly reversible process (isolable or independent reversible process) the ratio $\Delta S/s_1 - s_2$ will be always equal to unity.

It is a pleasure to acknowledge our indebtedness to Prof. D. S. Kothari, of Delhi University, for his encouragement and advice.

References.

- (1) E. H. Kennard, Proc. Nat. Aca. Sci. xviii. pp. 237-241 (1932).
- (2) D. S. Kothari and B. N. Singh, *Z. Astrophys.*, xv. p. 146 (1938).
- (3) B. N. Srivastava, Proc. Roy. Soc. A, clxxi. p. 480 (1939).

* This can be easily seen if we consider the approximate expression for pressure

$$p = \frac{2}{5} \frac{N^{5/3}}{V^{2/3}} \alpha.$$

We have $p_2/p_1 = (V_1/V_3)^{2/3}$ and we know⁽³⁾ that in the case of thermal transpiration, when the gas in both the chambers is degenerate (in the sense of Fermi-Dirac statistics), then in the equilibrium state, $p_1 = p_2$ to a first approximation. Hence $V_1 = V_3$.

LXXXVIII. *An Analysis of Hardness.*

By A. C. VIVIAN, D.Sc.*

[Received June 12, 1944.]

A CERTAIN quality or property of matter and of materials has been universally recognized and freely used in describing solid substances. According to this, opposition of varying degree is offered by "hardness" to penetration, denting, scratching, abrasion and wear. Hardness has been accepted by the majority as a simple fundamental property in its own right, as elementary or primary in fact; but it may be that the closest practical example of the thing in mind, the test characteristic of the conception, would prove on analysis to be mechanically complex and to display unsuspected features.

For, practical tests have ever been introduced to provide values or quantitative estimates to match conceptions of the properties. It not infrequently happens that such tests, undeniably to the point, will show the inadequacy or the inherent falsity of the very ideas from which the tests sprang. Nothing could have been more simple, apparently, than to have provided a figure for the property of "strength" by pulling a test piece in a machine until it broke; yet, the more elementary notion of strength which introduced the tensile test was proved inadequate by the examination of the test itself. For a long time the short-comings of the original notion blocked the road towards a better understanding of the behaviour of metals under load.

A mass of valuable information has been derived from the analysis of the tensile test; and any mechanical property not sufficiently well understood could be regarded as a potential source of supply of further knowledge of the behaviour of metals under load, and upon the relationships between the properties themselves. It seems to be of the greatest importance, therefore, to subject "hardness" to examination until its meaning is far more clear. Realizing how easily a conception handed on from the remote past might be critically erroneous, it would indeed be unwise to attempt to force preconceived ideas into the analysis. The object of the following is to accept the usual tests themselves, the actual examples after which the conception is straining, to show what hardness really is. If the mechanics of these tests can be clarified, the behaviour of materials undergoing those loadings known to elicit hardness will show what technical hardness is, how it should in fact be viewed whenever the necessity arises of framing a mental picture of it.

A superficial survey would show that the "scratch test," the "rebound tests," and the "indentation test" have been those commonly employed.

* Communicated by the Author.

Of these it is quite evident that the indentation tests are regarded as favourite criteria of the hardness of metals. Of these latter, the Brinell Hardness Test is the standard and the parent of similar tests which should be understood well enough if the Brinell test's meaning can be extracted by analysis.

To take the simplest Brinell test, it is carried out by loading a 0.01 mm. diameter ball (of very hard steel or of carbides), which is in contact with a smooth surface, with a load of 3,000 kilograms. This may take a few seconds to do, the load application often being made by the agency of a quick-acting oil pump until the full load is upon the ball and upon the material under it. Even after the full load has been applied in this manner the ball continues to sink into the material for a few seconds until it becomes practically stationary. The time allowed for the sinking-in is from 15 to 30 seconds; after which the load is taken off, and the diameter of the indentation is measured. From tables the Brinell Hardness Number (B.H.N.) corresponding to the diameter is now read, the "number" being just the stress in kilograms per square millimeter obtained by dividing the load (3,000) by the superficial area of the depression in contact with the ball at the conclusion of the allotted time of the test.

Preliminary Observations.

One line of investigation is suggested by the persistent and remarkable ratio—

$$\frac{\text{Ultimate tensile strength}}{\text{Brinell hardness number}} \text{ or } \frac{\text{U.T.S.}}{\text{B.H.N.}} = 0.22 \text{ approximately.}$$

For a large number of different alloys, notably the steels and duralumins of widely differing compositions, heat treatments and properties, the above ratio holds quite closely. On making allowance for the difference in the units employed in the numerator and denominator above, it would appear that the B.H.N. is 2.9 times as great a stress as the ultimate tensile stress. Very superficially viewed, such a relationship might appear reasonable: it is welcome enough for the opportunity it provides of approximating to the U.T.S. by means of the quick Brinell test. But, just because the real nature and inner meaning of the U.T.S. has become so well understood of late, this relationship is a most unwelcome puzzle to the discriminating, and will be described in the following analysis as the "mysterious" ratio. Examples of the constancy of this ratio are so well known to all that there seems no need to cite them.

The significance of the U.T.S. is now clear beyond any doubt. This nominal, unreal and incorrect stress value is merely incidental to a certain arbitrary manner of calculating the stress at maximum uniform extension; and yet the mysterious ratio refers to this erroneous stress specifically, and in no wise to the corresponding true stress. The ratio True Value of U.T.S. to the B.H.N. is certainly not a constant.

The nominal stress known as the U.T.S. would appear to have no significance at all where there can be no tension involving the shape-change to be expected in the ordinary tensile test specimen. The

shape of the nominal curve of the tensile test, and the value of every point on it including the U.T.S., depend on the contraction which takes place in ordinary tensile testing which does not occur in the indentation tests. Tensions are certainly to be expected in indentation tests, but none involving the contraction of area which would countenance the employment of the U.T.S. anywhere in the mechanism of indentation.

Possibly, however, the mysterious ratio is not one directly implying the *meaning* of the U.T.S. In this case, then, perhaps the ratio refers to the value of the U.T.S. on the true stress-strain curve. Such a point on the true curve would be reached by horizontally projecting the highest point of the nominal curve upon it. It would appear that the U.T.S. values on the true curve are represented by points roughly midway between the elastic limits and the regions at which the true curves straighten out again after their bend-over. Such points in the middle of the bold sweeps from maximum to minimum modulus might possibly be looked upon as the points of mean yield strength. These points cannot be far removed from having the values of the mean true stress between the origin of the curve and the true stress at maximum uniform extension; for, if $f = k.s^n$ be the logarithmic shape of the true curve, the relationship between the points and the true stresses in question must be $(1+n)/(1-s)$: n and s are both small and similar in value.

At one time it seemed possible and not unreasonable to draw this relationship into a theory of indentation hardness which regarded indentation as inducing the high stress of the metal flow first, which stress gradually drops as the area in contact with the ball increases. According to such a view, this point of mean stress might be significant as that point on the true stress-strain curve at which the test would terminate when the increasing area of contact and the decreasing stress between them called a halt to the descent of the load into the metal surface.

To express much the same idea in another way. In the tensile test, the U.T.S. is the (nominal) stress at which a rate of reduction of area overtakes a logarithmic rate of increase in the material's reactionary stress. No such crisis can occur in the usual compression test, but only in a compression test run backwards, as in the indentation tests possibly if they can be truly regarded as inducing the highest stresses first. Such a test would conclude automatically when a progressively increasing contact area overtook a progressively falling reactionary stress to oppose the applied load and the inertia of its descent. It will be observed that such an explanation is vague and unsubstantiated.

A Mathematical Approach.

It would be as well to recognize another possible implication of the mysterious ratio. If the possibility of a stress of the value of the U.T.S. but in kg.sq.mm. units be considered, then a new proposition is offered. Take the case of a 40-ton steel with a B.H.N. of 184:—

$$\frac{\text{U.T.S.}}{\text{C.H.N.}} = \frac{40 \text{ kg.sq.mm.}}{184 \text{ kg.sq.mm.}} = 0.22.$$

40 kg.sq.mm. is about 25 tons sq. in. Could the ratio refer to a B.H.N. relationship with the elastic limit or yield stress? If this relationship were the meaning of the ratio, the implications must be:—

(a) That there is some relationship between the U.T.S. and such an elastic limit or yield stress.

(b) That the conversion factor for kg.sq.mm. to tons sq. in., 1.575, happens to be the factor connecting the U.T.S. with the elastic limit or yield stress. That this might be the case as an approximation is suggested by Bauschinger's figures for the Bessemer carbon steels, for example, between 0.1 and 1.0 per cent. of carbon, the ratio between the U.T.S. and the elastic limit of these steels as rolled apparently only varies between 1.5 and 1.7.

Whatever this elastic limit or near-elastic point of stress actually proves to be, the mysterious ratio must mean that the B.H.N. is about 4.6 times that stress in the case of the steels and duralumins. The following paragraphs are intended to show how it comes about that the implications (a) and (b) are either justified or avoided, what near-elastic limit is $1/4.6$ of the B.H.N., and what "hardness" should then mean.

The case of a compression specimen, member, bar or disc of which its lateral strain is restricted in varying degrees has been worked out for elastic stressing and straining: it is well understood. As soon as this has been realized, and as soon as the striking similarity between indentation loading and the aforementioned case has been recognized, it becomes an irresistible and justifiable conviction that hardness manifestations of all kinds are attributable to the comparatively simple phenomena of local compression. It becomes impossible to attribute hardness to anything else but to the effect of the surrounding material's lateral restriction of the axial local compressive forces brought to bear either by a penetrating projectile, cutting tool, scratching particle of wind-born sand, or any other such agency. To deal with hardness manifestations as though they were peculiar in any respect is strange indeed when it is realized that the case in question is regularly treated in engineering mathematics which only need slight adaptation to explain hardness. Seeing that literally complete restriction of lateral strains must give the stress-strain curve of the axial loading the shape of the steep straight line of compressibility, it can naturally be said that hardness is some function of compressibility. On the other hand, compressibility is some function of the ordinary elasticity of the material, and it is not essential to bring compressibility into the picture.

Probably the simplest explanation is forthcoming from the consideration of the state of stress in a minute cubic element of material immediately under the centre of the Brinell ball of 5 mm. radius (R). Let this element be situated in the first zone of material exhibiting completely elastic reaction, that is, immediately below the last zone in depth exhibiting plastic reaction. Let the stress induced here by the applied compressive load in the test be f_x ; let this point be at Ymm. below the ball-centre; and let B.H.N./C be equal to f_x , C being a divisor

to convert the B.H.N. into the axial stress of at depth Y below ball-centre, or at $Y-5$ mm. below the surface of the indentation.

If there has never been anything but elastic reaction at this point Y , the elastic formulæ must apply. Upon the establishment of equilibrium at this little element of material, the lateral stresses indirectly induced by the potential lateral strains of f_x appear: these are f_y and f_y at right angles to one another and to f_x . In equilibrium, the reactions of f_x , f_y and f_y upon one another must be such that

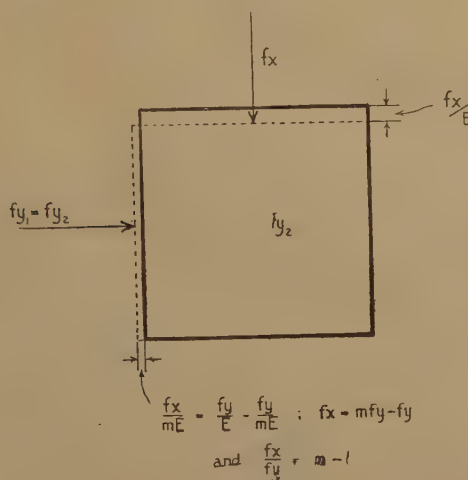
$$f_x - \frac{2f_y}{m} = f_y - \frac{f_y}{m},$$

whereby

$$\frac{f_x}{f_y} = m - 1,$$

m being Poisson's ratio (fig. 1).

Fig. 1.



At any value of Y the depth below the ball-centre, where purely elastic reactions according to Hooke's law are taking place, and where m can therefore be taken as the normal 3.3 for the steels, f_x must always be 2.3 times f_y .

The "stress difference" in such cases represents the *maximum* single stress in the material: "failure" takes place whenever this is equal to some yield of the material. It must be remarked, however, that failure in the sense of passing the limit of proportionality within which Hooke's law operates may only cause m to be very slightly reduced in value for purposes of the same equation as above. Failure in the ordinary sense of passing the elastic limit may need a still lower value of m in the equation, but the equation is still useful if the right m value can be found. Failure at a strong yield point may call for a much lower value of m . For, in perfect plasticity or fluidity as at considerable yield, m must come down to a value 2, whereby f_x and f_y would become equal, and the pressure in all directions would be the same as in a fluid. Still the formula applies whenever the right value for m can be found.

The stress difference $f_x - f_y$, at and below a Y which is below the limit of proportionality, may employ $m=3.3$. There are stress differences at smaller values of Y in zones at which the elastic limit and yield point are to be found: in these, particularly in the yield point zone, m will have lower values.

Let the word "failure" be used now for convenience in a non-committal sense to be applied to any stress $f_x - f_y$, thus:—

$$f_{\text{failure}} = f_f = f_x - f_y = f_x \cdot \frac{m-2}{m-1}.$$

Now, f_f (tons sq. in.) = f_f (kg.sq.mm.)/1.575.

Substituting B.H.N./C for f_x ,

$$f_f \text{ (tons sq. in.)} = \frac{\text{B.H.N.}}{C} \cdot \frac{m-2}{m-1} \cdot \frac{1}{1.575}.$$

Meanwhile,

$$\text{U.T.S. (tons sq. in.)} = f_f \text{ (tons sq. in.)} \times X,$$

where X is the conversion factor $\text{U.T.S.}/f_f$.

$$\text{Then, } \frac{\text{U.T.S.}}{\text{B.H.N.}} = \frac{1}{C} \cdot \frac{m-2}{m-1} \cdot \frac{X}{1.575}.$$

It now remains to discuss the factors C , X and m (fig. 2).

C is the divisor to convert the B.H.N. into the axial stress f_x at Y mm. below the ball-centre. Strain-etching of the cold work under the ball in the Brinell test on a number of carbon steels clearly enough shows the marked failure of the metal under the load to a depth of almost exactly 2 mm., as measured just below the ball-centre. Here, at 7 mm. below the ball-centre, is to be found the sharp line of demarcation of the zone of the yield point in all the low and medium carbon steels. Below this again, down to a depth of about 3 mm. (8 mm. from the ball-centre), are much fainter signs of cold work. It could, perhaps, be assumed that any purely elastic reaction must have taken place below the point at which $Y=8$ mm. (fig. 2).

The lines of compressive stress should be equally distributed over, and normal to the surface of, the depression; and, in the case of the lines immediately under the ball-centre, should be straight. This means that the stress at Y can be calculated as though the latter stress lines radiate from the ball-centre, though they do not do so in fact; and much evidence seems to favour the view that, for depths immediately below the ball-centre, the factor C is Y^2/R^2 .

The following calculation for a 40-ton steel (B.H.N. 184) will be found to justify that assumption moderately well. If all signs of cold work disappear below 8 mm., where C which is Y^2/R^2 is 2.56, it might be safe to take m as 3.3 there. Bauschinger has shown that the ratio U.T.S./elastic limit is 1.66 for such a steel: this is X . Hence in the case of this steel

$$\frac{\text{U.T.S.}}{\text{B.H.N.}} = \frac{1}{2.56} \cdot \frac{1.3}{2.3} \cdot \frac{1.66}{1.575} = 0.23 \text{ approximately.}$$

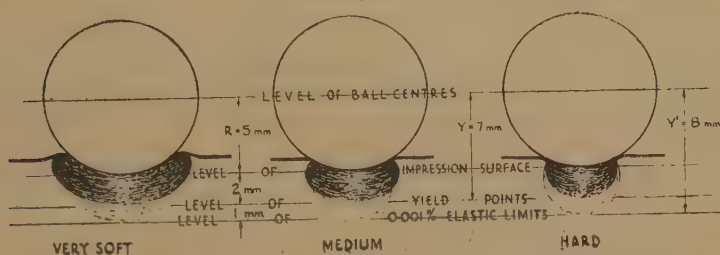
f_x (at $Y=8$ mm.) should have a value of $184/2.56$, which is 72 kg.sq.mm. ; f_y should therefore be $f_x/2.3$, which is 31.3 ; failure should therefore be at f_x-f_y or at 40.7 kg.sq.mm., or 25.8 tons sq. in. But Bauschinger's figure for the elastic limit of a 40-ton steel as rolled is about 24 tons sq. in. Evidently, then, 8 mm. is not quite the right depth below the ball-centre for Bauschinger's elastic limit. It will be found that, at 8.25 mm. depth for an elastic limit of 24 tons sq. in., the above calculation shows that C , or Y^2/R^2 , has a value of 2.72, and then

$$\frac{\text{U.T.S.}}{\text{B.H.N.}} = \frac{1}{2.72} \cdot \frac{1.3}{2.3} \cdot \frac{1.66}{1.575} = 0.22.$$

The observed $Y=8$ mm. is thus shown to be a slight underestimate.

One of the objects in taking the above example is to illustrate by a few figures how an identical relationship should apply also in the whole range of values of f_{failure} from the origin of the stress-strain curve right up to the point of considerable yield, say, up to 0.5 per cent. strain.

Fig. 2.



Strain-etched cold work under the 10-mm. Brinell ball.

R , the ball radius = 5 mm. ; Y , the depth of the yield point = 7 mm. ;
 Y' , the depth at elastic limit = 8 mm. (approx.).

In an aggregate of metal crystals like a steel there can hardly be any actual limit of proportionality at a definite point, but at the best only a region of stress beyond which the departure from Hooke's law and the falling off in the value of m becomes marked enough to attract attention. It will be seen that, for values of Y becoming progressively smaller at shallower depths, $1/C$, the first term of the equation, grows, while the second and third terms both diminish at the same time—the second term because the value of m drops, and the third term because the ratio X must naturally be dropping a little as the higher stress points are reached. Hence, there are compensating changes in the variation of the factors C , X and m which will probably be found to account for the application of the mysterious ratio at any and every point of the material immediately under the centre of the Brinell ball at which these factors can be estimated.

The following is a mere demonstration of the significance of the three combined terms—

$$\frac{1}{C} \cdot \frac{m-2}{m-1} \cdot \frac{X}{1.575}$$

as they relate to depths below the surface of the impression, and as they refer at the same time to values of f_f dropping from the stress-strain point of considerable plastic yield right down to the origin of the stress-strain curve. The assumption will be made that Bauschinger's 24 tons sq. in. elastic limit is at the point of permanent set of 0.00001, according to the recommendation of the International Congress of 1906. Up to the yield point, the curve shape of the steels (more particularly of those actually exhibiting the sudden yields) is steeper than beyond the yield point. If the yield were to continue, almost horizontal, as far as $s=1$, the stress of such a logarithmic curve ($f=k.s^n$) would be equal to k itself and to some such value as 27 which could be judged sufficiently closely for estimating the n value, of this early part of the whole stress-strain curve, at 0.01. From such values of k and n it can then be calculated that the standard (0.2 per cent. strain) yield point should be well below 26 tons sq. in. in this steel. If plastic strain be now chosen at such values as 0.0005, 0.001, 0.002, and 0.005, it can also be calculated from $f=27.s^{0.01}$ that the corresponding f_x-f_y , or f_f , stresses are respectively 25.0, 25.2, 25.4, and 25.6 tons sq. in.

A sufficient approximation to the nominal value of m ,—an apparent value of Poisson's ratio which can be described as " m ,"—at departures from Hooke's law at stress f_f by amounts of permanent set ϵ , can be arrived at by the use of a formula

$$"m" = \frac{(1+\epsilon)(1+f/E)-1}{\frac{1}{(1+\epsilon/2)} \cdot (1-f/mE)-1}.$$

These apparent values of Poisson's ratio have been calculated by the above formula for the chosen values of stress and strain; and the following table summarizes all such calculations and gives an indication of the manner in which the three factors C , X and m (also " m ") are presumed to vary and to compensate for each other as they vary.

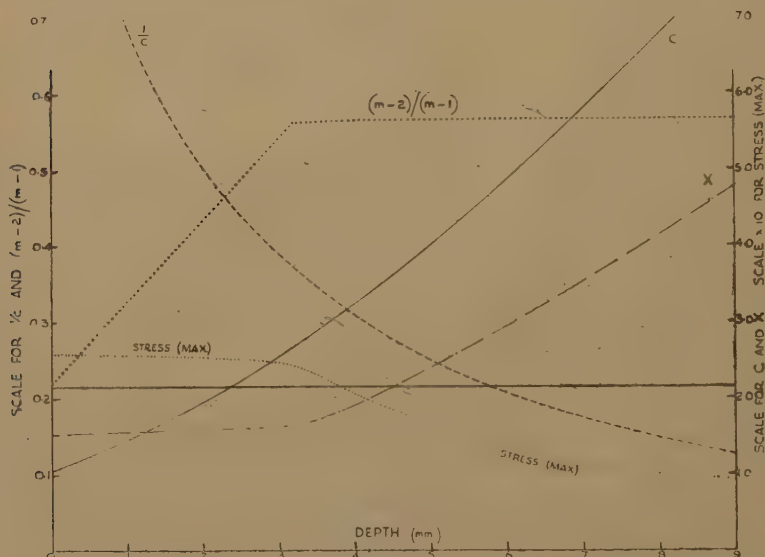
Since $1/C=R^2/Y^2=25/(\text{depth}+5)^2$, there is no difficulty in plotting the variables $1/C$, X , m and " m " on a base of depth in millimetres below the surface of the Brinell impression. This has been done in the following graph (fig. 3), from which the significance of a compensation between the terms becomes clearer still. It can be seen, for example, that while m is constant up to about 3 mm. depth, $1/C$ and X compensate each other. Then, as X levels out, the rapidly falling values of " m " compensate for $1/C$ rising as before.

The asterisk at C and $1/C$ in the table is intended to draw particular attention to the fact that these have been calculated on the assumption that the theory as to what is taking place is admissible: the table and graph prove very little or nothing, and are useless without experimental evidence in support or in the absence of good argument in favour.

The evidence forthcoming is restricted to observations from experimental trials that, on using a 10-mm. ball in the Brinell test, a number of steels show a strong line of the demarcation of the heavily cold-worked region almost exactly at $Y=7$ mm., just beneath the ball-centre,

irrespective of carbon content and strength. Now, the 0.2 per cent. yield-point of the 40-ton steel must be near 25 tons sq. in., and the completely independent computation of " m " assists the theory which has been advanced to place that yield-point at 25.3 tons sq. in. These steels also show signs of cold work down to about $Y=8$ mm., but the latter fade away so gradually that it is impossible by strain-etching to estimate where the 24 ton sq. in. (presumed at 0.001 per cent. strain) elastic limit may be in depth under the ball-centre. Had strain-etching been more effective, the weak signs must necessarily have extended further downwards; but, if it had been too effective it would have been equally inefficient as a guide to the point of 0.001 per cent. strain. It so happened that the rough estimate of the point of the elastic limit was not

Fig. 3.



very far out: it was at $Y=8$, instead of at $Y=8.25$ as the theory demands.

Hence, all that can be said about the experimental evidence is that it has approximately placed two highly important points by the employment of a theory which is, after all, just the mechanics of materials under three principal stresses, and a formula, $C=Y^2/R^2$, for a divisor to convert the B.H.N. into the axial stress at depth Y . The argument in favour of this formula has already been supplied: it is suggested that it should be taken on its merits, and not denounced because notable workers have sponsored other relationships between stress and depth which either do not apply to the particular case of the Brinell test, or not to the element immediately under the ball-centre.

If C has been correctly assessed, and if the mechanics of the three principal stresses f_x , f_y and f_z have been properly handled, there is no doubt that, at two depths below the ball-centre, the described relationship must exist between the three terms, which accounts for the

mysterious relationship between the U.T.S. and the B.H.N. If this is only a coincidence, it must surely be one even stronger than the ratio itself.

Observations on the Factor X.

The numerical relationships found to exist between the U.T.S. and the elastic limit or yield stress are themselves not without interest: they lie at the root of the mysterious ratio, and have hitherto been taken for granted as matters of casual fact.

It has already been mentioned that, according to Bauschinger, the whole range of Bessemer carbon steels as rolled shows a variation of X between 1.5 and 1.7 only, where X is the U.T.S./elastic limit ratio. But it is also well known that, amongst alloy steels and quench-hardened and tempered steels, the ratios between U.T.S. and yield-point can vary widely from about 1.5 to 1.1. Occasionally, figures are found which suggest that the U.T.S./elastic limit ratio of even these steels may be about 1.6 on the average; but it seems difficult to obtain definite evidence of the latter ratio for the alloy and heat-treated steels.

The nature of these relationships can be shown thus:—

Let $f = k \cdot s^n$ and $f' = k' \cdot s'^{n'}$ be the shapes of the first and the second parts of the stress-strain curves of all the alloys exhibiting merely temporary precipitation strain-strengthening during the first part of their curves. Now, in the steels, this phenomenon is sometimes accompanied by a notable break, or “jogged” yield, between the part of the curve affected by the precipitation and the remainder, and sometimes there is only a perfectly smooth transition from one curve-shape to the other.

Whenever it is possible to state a definite yield-point from automatic recording, it is possible to compute that the k value of the first curve-shape is only a little higher than the yield-point, and that s^n is practically unity at yield-point. Bearing this in mind for future use, it is necessary to observe that, if f_t be the true stress at maximum uniform extension in the tensile test, *i. e.*, the true value of the U.T.S., then at f_t the load, or $k' \cdot s'^{n'}/(1+s)$, is maximum. Differentiating and equating to zero gives

$$s = \frac{n}{1-n} \text{ at the strain of maximum uniform deformation (fig. 4).}$$

Load at $f_t = \text{U.T.S.} = k' \cdot s'^{n'}/(1+s)$; and substituting $n'/(1-n')$ for s —

$$\text{U.T.S.} = k' \cdot n'^{n'} \cdot (1-n')^{1-n'},$$

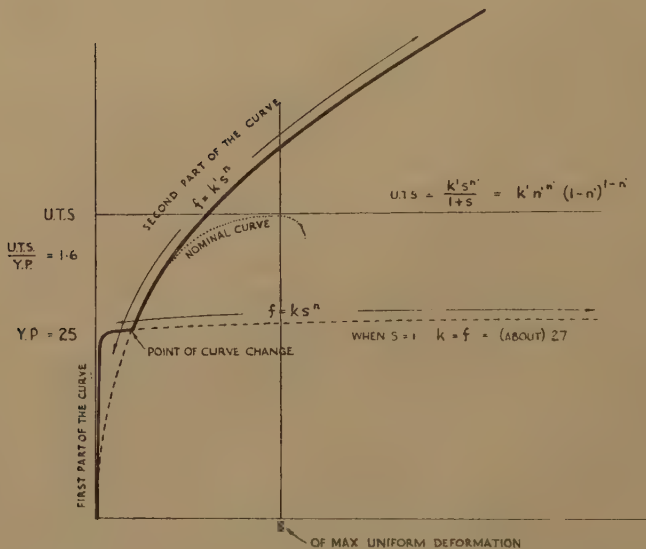
$$\frac{\text{U.T.S.}}{\text{Y.P.}} = \frac{k'}{k} \cdot n'^{n'} \cdot (1-n')^{1-n'}.$$

Between extreme and unusual values of s , the above terms containing n' vary between 0.5 and 1.0. Between commoner extreme values of s , they vary between 0.54 and 0.72. It is suggested that, at higher solution pressures of carbon in manufacture and heat treatment, *i. e.*, for the higher carbon steels (of normalized pearlitic structure), the ratio k'/k falls as the terms containing n' rise. In this way there would be a certain constancy in the ratio U.T.S./Y.P., at least in these steels. If this roughly constant ratio is about 1.6 as an average, the meaning is that

the ratio k'/k commonly varies between 3.0 and 2.2 as the strain of maximum uniform deformation varies between 0.3 and 0.1. If, in some steels and by some heat treatments, there is no tendency for a variation in k'/k , the U.T.S./Y.P. ratio might change considerably (as it is known to change) between 1.1 and 1.6 in the wide range of alloy and quench-hardened and tempered steels. Where there is little or no early precipitation effect in the early part of the curve, or where any precipitation effect continues unchanged throughout the whole curve, k' must equal k so that $k'/k=1$, and the ratio U.T.S./Y.P. must, for a 0.2 per cent. yield-point, be

$$\begin{aligned}\frac{\text{U.T.S.}}{\text{Y.P.}} &= \frac{k' \cdot n' n' \cdot (1-n')^{1-n'}}{k' \cdot (0.002)^{n'}} \\ &= (n'/0.002)^{n'} \cdot (1-n')^{1-n'}.\end{aligned}$$

Fig. 4.



When s varies between 0.3 and 0.1, this ratio would vary between 2.6 and 1.3. There is little doubt, however, that the alloy and quench-hardened steels likely to have no change in k throughout the curve are those stronger steels whose s values are far more likely to range between 0.15 and 0.05; in which cases the ratio would, in fact, vary between about 1.5 and 1.1 as found in practice with such steels.

Technical Hardness.

It would thus appear that hardness is a strength exhibited in loading by *local* compressive loads which, by the induction of lateral stresses in opposition to the development of the lateral strains, result in less axial strain (for any given degree of axial stress) than would be expected from the true stress-strain curve of single unidirectional loading. The

effect of such lateral stress induction is an appearance of greater strength in the material so loaded, since any axial stress induced under these conditions is no longer accompanied by a maximum shear stress of half its value, but by one of $(m-2)/(m-1)$ of half its value; from which it must appear that the axial stress induced by the load has been reduced by the factor, $(m-2)/(m-1)$, that the strain for a given stress is less than expected from the true stress-strain curve, and that the true stress-strain curve is steeper than it really is—*i. e.*, the material appears to be stronger in the proportion of $(m-1)/(m-2)$, or 1.77 times as strong in steels within the elastic range of stress.

Hardness will therefore depend upon strength proper and on Poisson's ratio, or rather on $(m-1)/(m-2)$, in elasticity. Strength itself beyond the elastic range depends on rate of strain-ageing, and nothing further need be said here concerning the latter essential feature of the stress-strain curve in plasticity. In plasticity, Poisson's ratio is succeeded by "*m*" values falling towards a limiting value of "*m*"=2, which metals do not attain even at their points of fracture, though they may do so at yield points for a period during which the curve lies horizontal.

In plasticity the material appears to be stronger than it really is in the proportion of $(\text{"m"}-1)/(\text{"m"}-2)$, depending on "*m*"; and in advanced plastic yield where, as at $Y=R$, "*m*" has been estimated as 2.28, the material should appear to be 4.6 times as strong as it is. This is also clearly reflected in the ratio $U.T.S./B.H.N.=0.22$ if this means $B.H.N./\text{elastic limit}=4.6$; for, obviously, the material in contact with the ball is standing a stress 4.6 times its elastic limit with impunity.

In an alternative view of questions of plastic deformation under hardness conditions of loading, "*m*" values may be discarded and ordinary *m* values effectively employed instead at the moment strain strengthening can be said sufficiently to have raised the elastic limit of the material to the stress point hitherto belonging to the plastic range. Strain strengthening, at least to the degree warranting the substitution of *m* for "*m*," must be extremely rapid in the steels; hence the elastic factor 1.77 may be employed without appreciable error upon the maximum true fracture stress of this steel, 65 tons sq. in., to procure the figure 115 tons sq. in. which represents the highest possible stress under hardness conditions of loading. It will be seen that this is 4.6 times a 25 ton sq. in. elastic limit, and is, of course, the $B.H.N.$ of 180 Kg. sq. mm.

The author would wish to conclude with offers of sincere thanks to those whose criticism has moulded his views, and whose co-operation has proved critically important in this analysis. In particular, he is indebted to Mr. G. H. Fisher for the patience and skill of the work of strain-etching under the Brinell impression; to Capitaine Kahan, of the French Air Force, for his assistance with the stress-difference criterion of failure; and to Dr. E. Orowan, for his help in the calculation of "*m*," to whom the author apologises for any misuse of that help.

LXXXIX. *The Physical Realizability of Electrical Networks having Prescribed Characteristics, with particular reference to those of the Probability Function Type.*

By F. F. ROBERTS, B.Sc. (Eng.), and J. C. SIMMONDS, Ph.D.*

[Received May 26, 1944.]

IN two earlier papers^{(1) (2)} the properties of exponential recurrent and probability function electrical pulses have been discussed, and it has been shown that interesting results are obtained when circuits having probability function responses are excited with probability function pulses. For example, if a pulse of this form is applied to a network also of this form, then the output wave has the same mathematical form as the applied pulse. However, in the earlier discussions, although a practical means of approximating the theoretical pulse wave-form was mentioned, no consideration was given to the practical possibility of obtaining the corresponding frequency response combined with the linear phase response tacitly assumed throughout the analysis. Discussion of this particular point would seem justified, for it may appear at first sight that the networks involved are not realizable in practice.

To show that the Probability Function Response is possible as an Approximation.

There are a number of different approaches to the general problem of the physical realizability of passive networks. Fundamentally, however, the condition for physical realizability is that no output shall be obtained from the network before the input signal is applied. If $H(\omega)$ represents the amplitude and θ the phase response at the frequency ω , then a necessary and sufficient condition for physical realizability can be shown to be⁽³⁾

$$\left. \begin{aligned} \frac{H(0)}{2} - \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) \sin(\omega t + \theta) d\omega = 0, \\ \text{for } t < 0. \end{aligned} \right\} \dots \dots \dots (1)$$

A slightly less general condition can be obtained from equation (1) by differentiation with respect to time :

$$\left. \begin{aligned} \int_{-\infty}^{\infty} H(\omega) \cos(\omega t + \theta) d\omega = 0, \\ \text{for } t < 0. \end{aligned} \right\} \dots \dots \dots (2)$$

* Communicated by the Authors.

In the present case the amplitude response is required to be

$$H(\omega) = Be^{-a\omega}, \quad (3)$$

and the phase response linear

$$\theta = b\omega. \quad (4)$$

Equation (2) then becomes

$$\int_{-\infty}^{\infty} Be^{-a\omega} \cos \omega(t+b) d\omega = 0, \quad \text{for } t < 0 \quad (5)$$

if the response functions are mutually consistent. The integral in the left of equation (5) is obviously not zero for all values of $t < 0$, thus showing that a network having the required response is impracticable*.

The same conclusion is also reached if the following attenuation/phase relation⁽⁴⁾, given by H. W. Bode, is applied to the problem. This attenuation/phase relation states that the phase angle β_c at any angular frequency ω_c is related to the attenuation by the equation

$$\beta_c = \frac{2\omega_c}{\pi} \int_0^{\infty} \frac{A - A_c}{\omega^2 - \omega_c^2} d\omega, \quad (6)$$

where A is the attenuation in nepers corresponding to ω . In the case under investigation.

$$A = \log(Be^{a\omega}) = \log B + a\omega, \quad (7)$$

so that

$$\beta_c = \frac{2a\omega_c}{\pi} [\omega]_0^{\infty}, \quad (8)$$

which shows that the required response cannot be obtained in practice. However, if the attenuation is given by equation (7) up to some frequency ω_1 , say, and at all higher frequencies the attenuation remains constant at A_1 , then from equation (6) it is seen that

$$\beta_c = \frac{2\omega_c}{\pi} \left\{ a\omega_1 + \frac{a(\omega_1^2 - \omega_c^2)}{2\omega_c} \left[\log \left(\frac{\omega - \omega_c}{\omega + \omega_c} \right) \right]_{\omega_1}^{\infty} \right\}. \quad (9)$$

When ω_1 is very large compared with ω_c , then

$$\beta_c \rightarrow \frac{4a\omega_c\omega_1}{\pi} \quad (10)$$

to a closer degree of approximation as ω_1 increases relative to ω_c . That is to say, in this case the phase law is linear in the limit, with a slope which tends to infinity as ω_c , the frequency beyond which the attenuation is constant, increases indefinitely.

This result shows that a probability function amplitude response and a linear phase response are physically consistent, provided that these responses are restricted to a finite frequency range (however large).

* It is interesting to note that any amplitude function $H(\omega)$ is possible if $\theta = \frac{\pi}{2} + b\omega$, since $H(\omega)$ is essentially an even function of frequency. A discussion of networks having this value of phase constant has been given by M. Levy⁽⁵⁾.

A Possible Method of Designing a Network with a Probability Function Type Response.

A possible method of realizing networks with responses of the type under discussion will now be given. The method depends upon the infinite product formulation of the exponential function

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \quad \dots \dots \dots (11)$$

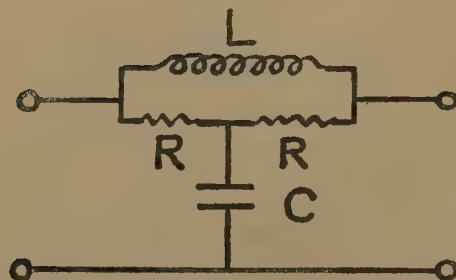
Making use of equation (11) the probability function may be written

$$e^{-a\omega^2} = \lim_{n \rightarrow \infty} \left(1 + \frac{a\omega^2}{n}\right)^{-n} \quad \dots \dots \dots (12)$$

By choosing for n a sufficiently large number, r say, the following approximation is obtained :

$$e^{-a\omega^2} \doteq \frac{1}{\left(1 + \frac{a\omega^2}{r}\right)^r} \quad \dots \dots \dots (13)$$

Fig. 1.



Now such a characteristic would obviously be given by a tandem connection of r similar networks each having an amplitude/frequency relation given by

$$H(\omega) = \left(1 + \frac{a\omega^2}{r}\right)^{-1} \quad \dots \dots \dots (14)$$

This type of characteristic is precisely that obtained from two equal bridged-T sections of the form shown in fig. 1. The phase constant of one such section is

$$\theta(\omega) = \tan^{-1} \omega \sqrt{\frac{a}{r}}, \quad \dots \dots \dots (15)$$

or in terms of ω_0 , the frequency at which the output is $1/\sqrt{2}$ times the input,

$$\theta(\omega) = \tan^{-1} \frac{\omega}{\omega_0} \quad \dots \dots \dots (16)$$

Thus, for values of ω such that $\omega \ll \omega_0$, or in other words, so long as the attenuation expected from each section is small, the phase constant is sensibly linear,

From equation (14), and using the parameter ω_0 , it is seen that the response of a single section of this type is given by

$$H(\omega) = \left\{ 1 - \frac{\omega}{\omega_0} \right\}^{-1} \dots \dots \dots (17)$$

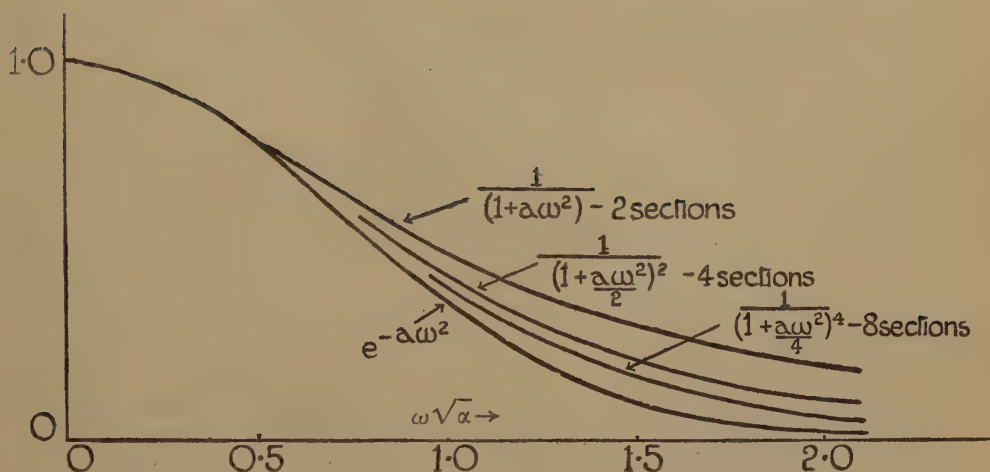
The component values of such networks are most conveniently expressed in terms of ω_0 and the characteristic impedance R_0 , which has a constant resistive value at all frequencies. In terms of these two quantities they are

$$\left. \begin{aligned} R &= R_0, \\ L &= R_0/\omega_0 = R_0 \sqrt{a/r}, \\ C &= 1/\omega_0 R_0 = \sqrt{a/r}/R_0, \end{aligned} \right\} \dots \dots \dots (18)$$

since

$$\omega_0 = \sqrt{r/a} \dots \dots \dots (19)$$

Fig. 2.



The amplitude response of two, four, eight and an infinity of simple network sections are shown in fig. 2, and the phase response in fig. 3. From these figures it is seen that a tolerably good approximation can be obtained with a value of r equal to 4, that is, with eight simple sections in the composite network, and that the degree of approximation is closest in the region of low attenuation, where errors are more likely to have disturbing effects on the transient response.

The Simulation of a High-Pass Response of Probability Form.

In a previous paper it was suggested that a high-pass response could be represented by the function

$$H(\omega) = B_1(1 - e^{-a_1\omega^2}), \dots \dots \dots (20)$$

and it was assumed that an amplitude response of this form is consistent with a linear phase response. It remains to show, therefore, that such

a response is physically possible and how it can be obtained in practice. Now the complex response of a composite network of the type dealt with in the previous section is of the form

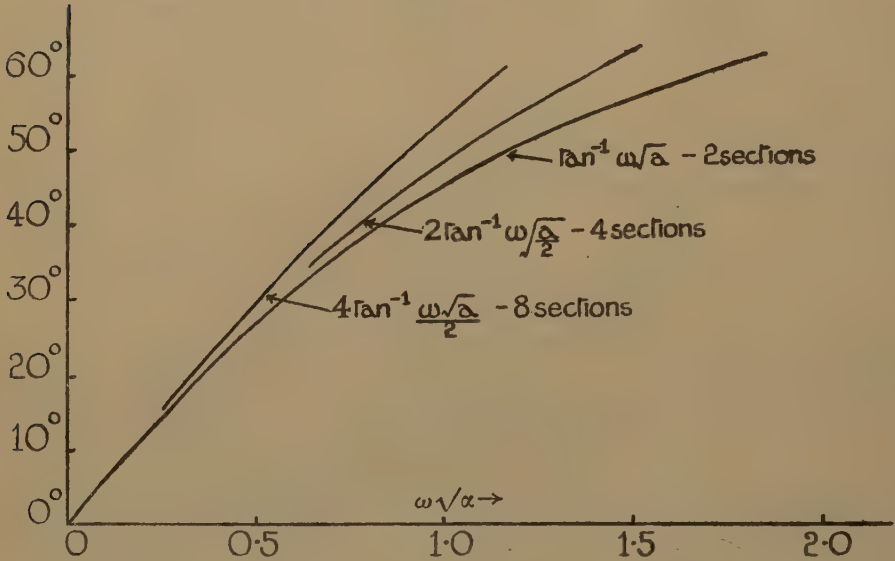
$$h(\omega)=B_1e^{-a_1\omega^2+jb_1\omega} \dots \dots \dots (21)$$

It is obvious that a network can be produced for which

$$h(\omega)=B_1e^{jb_1\omega}, \dots \dots \dots (22)$$

since this is the response of a delay network with fixed attenuation. If, then, two networks characterized by equations (21) and (22) are connected in parallel on the input side and in series on the output side

Fig. 3.



(through isolating transformers or valve stages if necessary), it is obvious that a complex response for the new network given by

$$h(\omega)=B_1e^{jb_1\omega}(1-e^{-a_1\omega^2}) \dots \dots \dots (23)$$

can be obtained. This response has linear phase and the amplitude response of equation (20).

Conclusions.

It has been shown how probability function type amplitude with linear phase response networks can be approximated in practice. This contribution completes the basic material required for a more detailed theoretical and practical investigation of the properties of pulse wave-forms and networks having probability function characteristics, and for the analysis of problems in which both pulses and responses of this type are involved.

The writers would like to thank the Engineer-in-Chief of the British Post Office for permission to publish this series of papers, and Dr. R. F. J. Jarvis for some helpful criticism.

References.

- (1) F. F. Roberts and J. C. Simmonds, "Some Properties of a Special Type of Electrical Pulse," *Phil Mag.*, Dec. 1943.
- (2) F. F. Roberts and J. C. Simmonds, "Some Further Properties of Exponential and Probability Function Pulse Wave-forms," *Phil. Mag.*, July, 1944.
- (3) E. A. Guillemin, 'Communication Networks' Vol. II., Chapter XI.
- (4) H. W. Bode "Improvements in or Relating to Thermionic Amplifiers." British Patent Specification 514, 567.
- (5) M. Levy, "The Impulse Response of Electrical Networks, with special reference to the Use of Artificial Lines in Network Design." *J. I. E. E.*, vol. xc. part iii., no. 12.

XC. A Note on the Minimum Radius for Degenerate Stellar Masses.

By D. S. KOTHARI and F. C. AULUCK
(University of Delhi) *.

[Received August 27, 1943.]

1. The Fermi-Dirac Statistics was first applied to the study of white dwarf stars by Fowler ⁽¹⁾ in a fundamental paper, and since then the theory of stellar configurations composed of degenerate matter has been extensively studied by Stoner ⁽²⁾, Milne, and Chandrasekhar ⁽³⁾. Chandrasekhar has worked out in complete detail the mass-radius relation in the case of complete degeneracy, taking account of the effect of relativistic mechanics. An important result of these investigations has been to show that for a mass larger than a critical mass M_c the radius R of the configuration vanishes. In this note we shall show that the theory of degenerate stellar masses composed of relativistic degenerate gas when carried through properly predicts a non-zero radius for any mass. The radius attains a minimum value for a mass which is about ten times larger than the Chandrasekhar-Stoner critical mass. For the purpose in view a crude treatment will suffice. The density of the configuration is assumed to be uniform.

If W denotes the gravitational potential energy of the configuration, and K the *residual* or *zero-point* kinetic energy of the degenerate matter composing it, then, we have as usual

$$W = -\frac{3}{5} \frac{GM^2}{R}, \quad \dots \dots \dots (1) \dagger$$

$$\text{and} \quad K = \frac{32\pi^2(mc)^5}{9mh^3} f(x)R^3, \quad \dots \dots \dots (2)$$

* Communicated by the Authors.

† The state of zero potential energy is taken to be that when the particles constituting the configuration are all dispersed to infinity.

where
$$f(x) = \frac{3}{8}x(1+x^2)^{1/2}(1+2x^2) - x^3 - \frac{3}{8}\log\{x+(1+x^2)^{1/2}\}, \quad (3)$$

$$x = \frac{h}{mc} \left(\frac{3n}{8\pi} \right)^{1/3}, \quad (4)$$

n is the electron concentration, m the mass of the electron, and the other symbols have their usual meaning. If ρ denotes the density and μ the mean molecular weight per free electron, then, we have

$$n = \frac{\rho}{\mu H},$$

and
$$x = \frac{h}{mcR} \left(\frac{9M}{32\pi^2\mu H} \right)^{1/3}, \quad (5)$$

where H is the mass of the proton. The mass-radius relation for the configuration is immediately obtained by applying the *Virial Theorem* *

$$2K + W = 0. \quad (6)$$

We have, after substituting for W and K ,

$$\frac{M}{\odot} = \frac{\alpha}{\mu^2} \left\{ \frac{f(x)}{x^4} \right\}^{3/2}; \quad \alpha = \frac{3}{2\pi H^2 \odot} \left(\frac{5ch}{3G} \right)^{3/2} \doteq 30, \quad (7)$$

where \odot is the mass of the Sun. The mass-radius relation is obtained by substituting for x from equation (5). The critical mass M_c is given by

$$\begin{aligned} \frac{M_c}{\odot} &= \frac{\alpha}{\mu^2} \lim_{x \rightarrow \infty} \left\{ \frac{f(x)}{x^4} \right\}^{3/2} \\ &= \frac{\alpha}{\mu^2} \left(\frac{3}{4} \right)^{3/2} \doteq \frac{19.5}{\mu^2}. \end{aligned} \quad (8)$$

2. In the above derivation of the mass-radius relation and the critical mass no distinction has been made between the mass of the configuration, when its particles are dispersed to infinity, and the mass when it has condensed into an equilibrium configuration. If M represents the mass of the configuration when the particles are dispersed to infinity, then the total energy (including the rest-mass energy) of the configuration in the equilibrium state is

$$E = Mc^2 + K + W = Mc^2 - K, \quad (9)$$

and thus the effective mass of the equilibrium configuration \bar{M} is

$$\bar{M} = E/c^2. \quad (10)$$

Ordinarily it is not necessary to distinguish between M and \bar{M} , but it is precisely this distinction which imposes a lower limit on the radius. The result $R \rightarrow 0$ for $M \geq M_c$ is obtained only when we ignore the distinction between M and \bar{M} . Using \bar{M} instead of M in the expression for the

* Stoner, instead of using the *Virial Theorem*, has applied the condition

$$\frac{d}{dR} (W + K) = 0.$$

gravitational potential energy, and noting that the number of electrons in the equilibrium state is the same as for the dispersed state, that is equal to $M/\mu H$, we have after applying the *Virial Theorem*

$$\frac{M}{\odot} = \frac{\alpha}{\mu^2} \frac{\{f(x)\}^{3/2}}{\left\{x^2 - \frac{m}{\mu H} \frac{f(x)}{x}\right\}^3}, \quad (11)$$

where $x = \frac{h}{mcR} \left(\frac{9M}{32\pi\mu H} \right)^{1/2}$, $\alpha = \frac{3}{2\pi H^2 \odot} \left(\frac{5c\hbar}{3G} \right)^{3/2} \doteq 30$. (12)

The relation between M and \bar{M} is easily found to be

$$\frac{\bar{M}}{M} = 1 - \frac{m}{\mu H} \frac{f(x)}{x^3}. \quad (13)$$

For the limiting case $x \rightarrow 0$ the above equations reduce to

$$\bar{M} = M, \quad (14)$$

$$R = \left(\frac{9\pi}{4} \right)^{2/3} \frac{L_0}{\mu^{5/3}} \left(\frac{\odot}{M} \right)^{1/3}, \quad (15)$$

where $L_0 = \frac{\hbar^2}{4\pi^2 m H^{5/2} G^{1/3}} \doteq 6.18 \times 10^8 \text{ cm}$.

As can be seen from equation (11), x cannot exceed the value $\frac{4\mu H}{3m}$. When x approaches this value equations (13) and (12) reduce to

$$\bar{M} = \left(\frac{3}{4} \right)^{1/2} \frac{(\alpha \odot)^{1/3}}{\mu^{2/3}} M^{2/3}, \quad (16)$$

$$R = \frac{\frac{3}{4} \left(\frac{9\odot}{32\pi^2} \right)^{1/3} \frac{h}{c(\mu H)^{4/3}} \left(\frac{M}{\odot} \right)^{2/3}}{\left(\frac{M}{\odot} \right)^{1/3} - \left(\frac{3}{4} \right)^{1/2} \frac{\alpha^{1/3}}{\mu^{2/3}}}, \quad (17)$$

or $R \sim \frac{3}{4} \left(\frac{9\odot}{32\pi^2} \right)^{1/3} \frac{h}{c(\mu H)^{4/3}} \left(\frac{M}{\odot} \right)^{1/3}$ as $M \rightarrow \infty$. (17 a)

We see, therefore, that as the mass increases, the radius of the configuration at first decreases, then attains a minimum value and finally increases with increasing mass. Differentiating equation (17) we obtain for the minimum radius, the value

$$R = \frac{9}{4\pi} \left(\frac{5}{4} \right)^{1/2} \frac{h^{3/2}}{c^{1/2} G^{1/2} (\mu H)^2} \doteq \frac{3.5 \times 10^6}{\mu^2} \text{ cm.}, \quad (18)$$

and the mass corresponding to this radius is

$$M_0 = 2\bar{M}_0 = 8 \left(\frac{3}{4} \right)^{3/2} \frac{\alpha \odot}{\mu^2}. \quad (19)$$

It may be noted that M_0 is eight times the Chandrasekhar-Stoner critical mass \bar{M}_c . The density of the stellar matter for a mass in the neighbourhood of and exceeding M_0 is of the order of nuclear density ($10^{15} \text{ gm./cm.}^3$).

All the same, it is of interest to note that the theory of degenerate stars does not lead to a vanishing radius for a mass exceeding a certain critical mass, provided account is taken of the change of mass due to loss of energy accompanying the condensation of the material.

References.

- (1) Fowler, Monthly Notices, R.A.S. lxxxvii. p. 114 (1926).
- (2) Stoner, Phil. Mag. ix. p. 944 (1930).
- (3) Chandrasekhar, "An Introduction to the Study of Stellar Structure" (1939), where other references are given.

XCI. Notice respecting New Book.

The Simple Calculation of Electrical Transients. By G. W. CARTER. [Pp. 120.] (Cambridge University Press.) Price 8s. 6d. net.

THIS small book has been written in order to provide electrical engineers with an elementary guide to the calculation of the currents and voltages in electrical networks under transient conditions; such as occur, for example, in the period of transition from one steady state to another, immediately following the closing or opening of a switch. The aim has clearly been to introduce Heaviside's operational method to engineers who have no extensive mathematical equipment, and who will be content to judge the soundness of the method by the extent to which the results agree with their practical experience of circuits rather than by purely mathematical considerations. This is not to imply that the mathematical groundwork is inadequate for the purpose. The reader is not encouraged to proceed by rule of thumb until he has first laid down a foundation for it, but he is never in any danger of forgetting that he is solving electrical problems rather than mathematical ones, and that the author's object is to enable him to do this intelligently, but with the least possible mathematical effort. The scope of the book is limited to circuits capable of being represented by combinations of a finite number of ordinary resistances, capacitances and inductances, self and mutual; that is to say, it stops short of non-linear elements and distributed circuits. One of its most useful features is a collection of examples, obviously based on actual practice, showing what a very wide range of problems in power engineering can be handled by so comparatively simple a technique; *e.g.*, switching problems, the effects of lightning discharges, and problems concerning resonance, damping and the stability of circuits. The treatment is always clear and very well adapted for the type of reader for whom it is intended. The book can be recommended to engineers and experimental physicists as a very useful introduction to the subject.

L. H.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XCII. *Energy Distribution in the Spectrum of a Frequency Modulated Wave.*—Part I.

By A. S. GLADWIN, B.Sc., A.R.T.C., A.M.I.E.E.

(Communication from the Staff of the Research Laboratories of
The General Electric Co., Ltd., England *.)

[Received August 14, 1944.]

Summary.

A method is derived of calculating the distribution of energy in the frequency spectrum of a frequency modulated wave when the distribution of energy in the frequency spectrum of the modulating wave is known, and the modulating wave consists of a very large number of sinusoidal components of unrelated frequencies and small amplitudes.

The spectra for two particular cases of frequency modulation by telephonic signals are calculated.

CONTENTS.

	Page
1. Introduction	787
2. List of symbols	788
3. The modulating wave	789
4. A general expression for the sideband energy distribution ..	790
5. The mean-squared value of the carrier wave	792
6. Energy distribution in the first order sideband spectrum....	793
7. Energy distribution in the second order sideband spectrum..	794
8. Higher order spectra	797
9. Modulation by telephonic signals.....	797

1. *Introduction.*

MOST of the published theoretical work on frequency modulation has dealt only with the case of a single frequency modulating wave. In the treatment of amplitude modulation and low-frequency problems there is some slight justification for this method, but in frequency modulation the single-frequency analysis gives little information concerning the more complicated cases which occur in practice.

Of the few analyses which have dealt with the case of frequency modulation by a multi-tone signal the most notable is that of Cherry and Rivlin ⁽¹⁾, in which expressions for the amplitudes and phases of all the sidebands are derived on the assumption that the modulating wave consists of a number of sinusoidal components. Sidebands are produced

* Communicated by C. C. Paterson, F.R.S.

corresponding in their frequency spacing from the carrier wave, not only to the frequencies in the modulating wave, as in amplitude modulation, but also to all the harmonics and to all the possible combinations of the sums and differences of all the harmonics of the frequencies in the modulating wave.

When the number of components in the modulating signal is large, as it is in most cases, it is not practicable to evaluate the amplitude of every sideband, and indeed the amplitude of a particular sideband is of little interest. What is usually required is the distribution of energy or power in the spectrum of the frequency modulated wave.

If the modulating wave consists of a large number of components the mean-squared value of the components per unit of band-width near to any given frequency is simply the sum of the mean squared values of the individual components lying within this unit range, and can, with sufficient precision, be represented as a continuous function of frequency.

If, in addition, the frequencies of the components of the modulating wave are unrelated, the frequency spacings between adjacent sidebands are also unrelated, and the sideband energy distribution or mean-squared value per unit of band-width can also be represented as a continuous function of frequency. In the following sections it will be shown how this function may be calculated when the function representing the energy distribution in the modulating signal is known.

2. List of Symbols.

a_m	= Amplitude of any sinusoidal component in the modulating wave.
C	= Crest factor = peak value/R.M.S. value.
D	= Deviation ratio = maximum frequency deviation / highest modulating frequency.
$f(\omega_m)$	= Energy distribution function for modulating wave.
$g_n(D, \omega_s)$	= Energy distribution function for the n th order sideband spectrum.
K	= Function of the modulating wave = $\int_{\omega_1}^{\omega_2} (\omega_s^2 / \omega_m^2) f(\omega_m) d\omega_m$.
J_n	= Bessel function of the first kind of order n .
P_o	= Mean-squared value of the carrier wave relative to the mean-squared value of the whole frequency modulated wave.
ω_1	= Lowest of the frequencies of the components in the modulating wave.
ω_o	= Carrier wave frequency.
ω_m	= Frequency of any component in the modulating wave.
ω_2	= Highest of the frequencies of the components in the modulating wave.
ω_s	= Difference in frequency between any sideband and the carrier wave.

- ω_{dc} = Instantaneous frequency deviation of a frequency modulated wave from the carrier wave frequency.
 $\Delta\omega_c$ = Maximum frequency deviation of a frequency modulated wave from the carrier wave frequency.
 ϕ_m = Phase of any component in the modulating wave.

3. The Modulating Wave.

The modulating wave is supposed to consist of a very large number of components of unrelated frequencies occupying a finite band of frequencies from the lower limit ω_1 to the upper limit ω_q . It is also supposed that the amplitude of each component is small compared with the amplitude of the complete wave. Denoting the modulating wave by S,

$$S = \sum_{m=1}^q a_m \cos(\omega_m t + \phi_m). \quad (3.1)$$

The energy density at a frequency ω_m may be defined in terms of mean-squared values as follows:—The mean-squared value of that part of the modulating wave lying within any frequency range of unit band-width is simply the sum of the mean-squared values of all the components which lie within this unit range of frequency. The energy density at a frequency ω_m is the mean value of this quantity in the neighbourhood of the frequency ω_m , and may be denoted by $f(\omega_m)$, which, for all practical purposes, can be regarded as a continuous function of ω_m . It is also convenient to make the total mean-squared value of the wave equal to unity.

$$\therefore \int_{\omega_1}^{\omega_q} f(\omega_m) d\omega_m = \sum_{m=1}^q \frac{a_m^2}{2} = 1. \quad (3.2)$$

If $\phi(\omega_m)$ is another function of ω_m it is obvious that

$$\int_{\omega_a}^{\omega_b} \phi(\omega_m) f(\omega_m) d\omega_m = \sum_{m=a}^b \phi(\omega_m) \frac{a_m^2}{2}. \quad (3.3)$$

Another property of the modulating wave which will be used in the analysis is the crest factor, denoted by C, which is defined as the ratio of peak to R.M.S. values. There is no difficulty in defining the peak value of a simple wave-form, but for complex waves consisting of a large number of components of unrelated frequencies the peak value becomes a somewhat indefinite quantity. In such cases the peak value may be defined as the value which is exceeded in either the positive or negative direction, or both, for a given fraction of the total time, *e. g.*, 10, 1, 0.1 per cent. This is an arbitrary, but none the less useful definition. Clearly the method of defining the peak value must be stated when giving a value for the crest factor. The R.M.S. value, too, depends on the time interval over which it is measured. In this paper, it is supposed that the interval of measurement is such that no change in the R.M.S. values takes place when the interval is increased.

Since the mean-squared value of the modulating wave has been taken as unity, the peak value is equal to the crest factor C.

4. A General Expression for the Sideband Energy Distribution.

Let the modulating wave, given by (3.1), be used to frequency modulate a carrier wave, $\sin \omega_c t$; and let the instantaneous frequency deviation of the modulated wave from the value ω_c be ω_{dc} .

Then
$$\omega_{dc} = k \sum_{m=1}^q \alpha_m \cos(\omega_m t + \phi_m),$$

where k is a constant.

Let the peak frequency deviation (the peak being defined after the manner of Section 3) be $\Delta \omega_c$. The peak value of the modulating wave is C,

$$\begin{aligned} \therefore \quad \Delta \omega_c &= kC. \\ \therefore \quad \omega_{dc} &= \frac{\Delta \omega_c}{C} \sum_{m=1}^q \alpha_m \cos(\omega_m t + \phi_m). \quad \dots \quad (4.1) \end{aligned}$$

The ratio of the maximum frequency deviation to the highest frequency in the modulating wave is known as the deviation ratio and is denoted by D.

$$\therefore \quad D = \frac{\Delta \omega_c}{\omega_q}.$$

Substituting this in (4.1), the expression for the frequency modulated wave can be written as

$$\sin \left[\omega_c t + \frac{D}{C} \sum_{m=1}^q \alpha_m \frac{\omega_q}{\omega_m} \sin(\omega_m t + \phi_m) \right]. \quad \dots \quad (4.2)$$

Cherry and Rivlin⁽¹⁾ have shown that in the spectrum of a frequency modulated wave given by an expression similar to (4.2), the amplitude of the sideband of frequency

$$\begin{aligned} &\omega_c + \sum_{m=1}^q \alpha_m \omega_m \\ \text{is} \quad &\prod_{m=1}^q J_{\alpha_m} \left(\alpha_m \frac{D}{C} \cdot \frac{\omega_q}{\omega_m} \right), \quad \dots \quad (4.3) \end{aligned}$$

where J_{α_m} is the Bessel function of the first kind of order α_m , and α_m can have any positive or negative integral value, or the value 0.

In Section 3 it was postulated that the amplitude α_m of any component in the modulating wave should be very small compared with the amplitude of the complete wave, viz. C. It will also be supposed that the quantity $\alpha_m \frac{D}{C} \cdot \frac{\omega_q}{\omega_m}$ is also small compared with unity for all the relevant values of the variables.

Since the argument of J_{α_m} is in every case small, it is permissible to use an approximation for $J_n(x)$ (which becomes in the limit an identity)

$$\begin{aligned} J_n(x) &\doteq \left(\frac{x}{2}\right)^n \left[1 - \left(\frac{x}{2}\right)^2 / (n+1)\right] / n! \\ &\doteq \left(\frac{x}{2}\right)^n [\exp\{-x^2/4(n+1)\}] / n!. \quad \dots \quad (4.4) \end{aligned}$$

Substituting this approximation in expression (4.3), the expression becomes

$$\begin{aligned} &\prod_{m=1}^q \left[\left(\frac{\alpha_m}{2} \cdot \frac{D}{C} \cdot \frac{\omega_q}{\omega_m} \right)^{\alpha_m} \left(\exp \left\{ \frac{-\alpha_m^2}{4(\alpha_m+1)} \cdot \frac{D^2}{C^2} \cdot \frac{\omega_q^2}{\omega_m^2} \right\} \right) / \alpha_m! \right] \\ &= \left[\exp \left\{ -\frac{D^2}{C^2} \sum_{m=1}^q \frac{\alpha_m^2}{4(\alpha_m+1)} \cdot \frac{\omega_q^2}{\omega_m^2} \right\} \right] \prod_{m=1}^q \left[\left(\frac{\alpha_m}{2} \cdot \frac{D}{C} \cdot \frac{\omega_q}{\omega_m} \right)^{\alpha_m} / \alpha_m! \right]. \quad (4.5) \end{aligned}$$

This expression is valid only for positive values of α_m . Since, however, $J_{-n}(x) = (-1)^n J_n(x)$, it is necessary only to substitute the numerical values of α_m into the expression in order to obtain the sideband amplitude when negative values of α_m are involved. The additional factor $(-1)^n$ may be ignored, since a negative sign merely indicates a change of 180° in the phase of the sideband.

It is also clear from the foregoing relations that for each sideband of frequency, say $\omega_c + \omega_s$, there exists a corresponding sideband of equal amplitude of frequency $\omega_c - \omega_s$. Thus the sideband energy distribution is symmetrical with respect to the carrier frequency. Only the sidebands of frequency greater than ω_c will be considered further. The difference in frequency between a sideband and the carrier will be denoted by ω_s .

Sidebands of frequency near to $\omega_c + \omega_s$ may be formed in a number of different ways, corresponding to the different values α_m may assume; for example, if ω_s lies between $n\omega_1$ and $n\omega_q$, a sideband of frequency $\omega_c + \omega_s$ is formed by a component of frequency ω_s/n in the modulating wave, and the sideband amplitude is found from expression (4.5) by putting $\alpha_m = n$ for $\omega_m = \omega_s/n$, and $\alpha_m = 0$ for all other frequencies. The simplest case is that in which $n=1$. Similarly, a sideband of frequency near to $\omega_c + \omega_s$ will be formed by any two components in the modulating wave, the sum or difference of any of whose harmonic frequencies is nearly ω_s . (The sum or difference will not be exactly ω_s , since the frequencies of the components are unrelated), and so on. The mean-squared value of all the sidebands of frequency close to $\omega_c + \omega_s$ is, therefore, the sum of the mean-squared values of the individual sidebands.

The sideband spectrum, which is formed by the components of the modulating wave taken one at a time, will be termed the first-order sideband spectrum, and the energy distribution in the spectrum will be specified in terms of the mean-squared value per unit of band-width in the same way as for the modulating wave, as described in Section 3. The sideband spectra formed by the components of the modulating

wave taken 2 . . . q at a time will be termed the 2nd . . . q th order spectra, and the complete expression for the sideband energy distribution will be the sum of the expressions for the energy distribution in the 1st, 2nd . . . q th spectra.

From expression (4.5) it can be seen that the energy distribution in the first-order spectrum is a function of D and ω_s , say $g_1(D, \omega_s)$. Similarly, the energy distributions in the 2nd . . . q th order spectra are functions $g_2(D, \omega_s)$. . . $g_q(D, \omega_s)$, and the complete expression for the sideband energy distribution is

$$g_1(D, \omega_s) + \dots + g_q(D, \omega_s). \quad (4.6)$$

In the following sections the first two terms of this series will be calculated. Before this is done, the mean-squared value of the carrier wave will be found.

5. The Mean-squared Value of the Carrier Wave.

If, in expression (4.5), α_m is zero for all values of m , the expression gives the amplitude of the component in the frequency-modulated wave of frequency ω_c , i. e., the carrier wave.

The carrier amplitude is

$$\begin{aligned} & \exp \left\{ -\frac{D^2}{4C^2} \sum_{m=1}^q \alpha_m^2 \frac{\omega_q^2}{\omega_m^2} \right\} \\ & = \exp \left\{ -\frac{D^2}{2C^2} \cdot \int_{\omega_1}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot f(\omega_m) d\omega_m \right\}. \quad \text{From (3.3).} \end{aligned}$$

The mean-squared value of the carrier relative to the mean-squared value of the whole wave is simply the square of this expression. Denoting this quantity by P_c ,

$$P_c = \exp \left\{ -\frac{D^2}{C^2} \cdot \int_{\omega_1}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot f(\omega_m) d\omega_m \right\} = \exp \left(\frac{-KD^2}{C^2} \right), \quad (5.1)$$

$$\text{where} \quad K = \int_{\omega_1}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot f(\omega_m) d\omega_m. \quad (5.2)$$

Since the mean-squared value of the complete frequency modulated wave relative to itself is unity, the total mean-squared value of the sidebands relative to the whole wave is

$$\begin{aligned} 1 - P_c &= P_c \left(\frac{1}{P_c} - 1 \right) \\ &= P_c \left\{ \exp \left(\frac{KD^2}{C^2} \right) - 1 \right\} \\ &= P_c \left\{ \frac{KD^2}{C^2} + \frac{K^2 D^4}{2! C^4} + \dots \right\}. \quad (5.3) \end{aligned}$$

It will be shown that the total mean-squared value of the sidebands

in the first order spectrum (including sidebands both above and below the carrier frequency) is equal to

$$P_c \frac{KD^2}{C^2},$$

and that the total mean-squared value of the sidebands in the second order spectrum is equal to

$$P_c \frac{K^2 D^4}{2! C^4}.$$

6. Energy Distribution in the First-Order Sideband Spectrum.

In this section the energy distribution in the sideband spectrum formed by taking the components of the modulating wave one at a time will be calculated.

The component in the modulating wave of frequency ω_m produces sidebands of frequency

$$\omega_c + \alpha_m \omega_m = \omega_c + \omega_s.$$

It has been pointed out in Section 4 that the energy spectrum of the frequency modulated wave is symmetrical with respect to the carrier frequency. Only positive values of α_m need therefore be considered. Taking first the case of $\alpha_m = 1$ ($\omega_m = \omega_s$) the amplitude of the sideband of frequency $\omega_c + \omega_s$ is, from (4.5) (putting $\alpha_m = 1$ for $\omega_m = \omega_s$ and $\alpha_m = 0$ for all other values of ω_m),

$$\frac{D}{2C} \cdot \frac{\omega_q}{\omega_s} \cdot a_s \exp \left[\frac{-D^2}{4C^2} \left\{ \left(\sum_{m=1}^q \frac{a_m^2 \omega_q^2}{\omega_m^2} \right) - \frac{a_s^2 \omega_q^2}{2\omega_s^2} \right\} \right]. \quad (6.1)$$

Since $\frac{a_s^2 \omega_q^2}{2\omega_s^2}$ is negligible compared with $\sum_{m=1}^q \frac{a_m^2 \omega_q^2}{\omega_m^2}$ the exponential function becomes

$$\exp \left[\frac{-D^2}{2C^2} \cdot \int_{\omega_1}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot f(\omega_m) d\omega_m \right].$$

From (6.1) and (5.1), the mean-squared value of the sideband relative to the whole wave is

$$\frac{P_c D^2}{4C^3} \cdot \frac{\omega_q^2}{\omega_s^2} \cdot a_s^2. \quad (6.2)$$

The mean-squared value of the sidebands per unit of band-width is the mean value of

$$\frac{P_c D^2}{4C^2} \cdot \sum_{\omega_s}^{\omega_s+1} a_s^2 \frac{\omega_q^2}{\omega_s^2}$$

in the neighbourhood of ω_s and

$$= \frac{P_c D^2}{2C^2} \cdot \frac{\omega_q^2}{\omega_s^2} \cdot f(\omega_s). \quad (6.3)$$

In the sideband spectra corresponding to values of $\alpha_m = n > 1$, the sidebands are spaced in frequency by n times the spacing of the corresponding components in the modulating wave, and the mean-squared value of the sideband relative to the whole wave is

$$P_c \left[\frac{D}{2C} \cdot \frac{n\omega_g}{\omega_s} a_{s/n} \right]^{2n} / (n!)^2.$$

($a_{s/n}$ = amplitude of the component of the modulating wave of frequency $\frac{\omega_s}{n}$).

The expression for the energy distribution corresponding to (6.3) is

$$P_c \frac{D^2}{2C^2} \cdot \frac{n\omega_g^2}{\omega_s^2} \cdot f\left(\frac{\omega_s}{n}\right) \left[\frac{D}{2C} \cdot \frac{n\omega_g}{\omega_s} \cdot a_{s/n} \right]^{2(n-1)} / (n!)^2. \quad (6.4)$$

Since $a_{s/n}$ is very small compared with 1, expression (6.4) is always negligible compared with expression (6.3).

So spectra corresponding to values of α_m greater than 1 may be neglected.

Equating (6.3) to the first term of (4.6),

$$\begin{aligned} g_1(D, \omega_s) &= P_c \frac{D^2}{2C^2} \cdot \frac{\omega_g^2}{\omega_s^2} \cdot f(\omega_s) \\ &= P_c \frac{D^2}{2C^2} F_1(\omega_s), \quad \dots \dots \dots (6.5) \end{aligned}$$

where

$$F_1(\omega_s) = \frac{\omega_g^2}{\omega_s^2} \cdot f(\omega_s). \quad \dots \dots \dots (6.6)$$

The total mean-squared value of the sidebands of frequency greater than ω_c in the first order spectrum is the integral of the energy density over the range $\omega_s = \omega_1$ to $\omega_s = \omega_g$.

From (6.3) this is

$$\begin{aligned} &P_c \frac{D^2}{2C^2} \int_{\omega_1}^{\omega_g} \frac{\omega_g^2}{\omega_s^2} \cdot f(\omega_s) d\omega_s \\ &= P_c \frac{KD^2}{2C^2}. \quad \text{From (5.2).} \end{aligned}$$

The total mean-squared value of the sidebands of frequency less than ω_c is also equal to this expression, so that the total mean-squared value of the sidebands in the first order spectrum is $P_c KD^2/C^2$. This is the first term in the series (5.3).

7. *Energy Distribution in the Second-Order Sideband Spectrum.*

In this section the energy distribution in the sideband spectrum formed by taking the components of the modulating wave two at a time will be calculated. In view of the findings of the last section only values of $\alpha_m = \pm 1$ will be considered,

If ω_m and ω_n are the frequencies of any two components in the modulating wave, the corresponding sideband frequencies are

$$\omega_c + \omega_m + \omega_n, \quad \omega_c + \omega_m - \omega_n, \quad \omega_c - \omega_m + \omega_n, \quad \omega_c - \omega_m - \omega_n.$$

If $\omega_m > \omega_n$, the frequencies of the first two sidebands are greater than ω_c , and the frequencies of the latter two are less than ω_c , and will therefore not be considered.

The sidebands of frequency $\omega_c + \omega_m + \omega_n = \omega_c + \omega_s$ will be dealt with first. These sidebands cover the frequency range $\omega_c + 2\omega_1$ to $\omega_c + 2\omega_q$. Obviously, sidebands of frequency near to $\omega_c + \omega_s$ will be formed by any two components of the modulating wave whose frequencies satisfy the relation $\omega_m + \omega_n \sim \omega_s$.

From expression (4.5) the amplitude of the sideband of frequency $\omega_c + \omega_s$ is

$$\frac{D^2}{4C^2} \cdot \frac{\omega_q}{\omega_m} \cdot \frac{\omega_q}{\omega_n} \cdot a_m a_n \exp \left[\frac{-D^2}{4C^2} \left\{ \left(\sum_{m=1}^q a_m^2 \frac{\omega_q^2}{\omega_m^2} \right) - \frac{a_m^2 \omega_q^2}{2\omega_m^2} - \frac{a_n^2 \omega_q^2}{2\omega_n^2} \right\} \right]. \quad (7.1)$$

The mean-squared value of the sideband relative to the whole wave is

$$P_c \frac{D^4}{16C^4} \cdot \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{\omega_n^2} \cdot a_m^2 a_n^2.$$

Keeping ω_m fixed and varying ω_n from ω_n to $\omega_n + 1$, the mean-squared value of the sidebands in the unit range ω_s to $\omega_s + 1$ is seen to be

$$\begin{aligned} & P_c \frac{D^4}{8C^4} \cdot \frac{\omega_q^2}{\omega_m^2} \cdot a_m^2 \sum_{\omega_n}^{\omega_n+1} \frac{\omega_q^2}{\omega_n^2} \cdot \frac{a_n^2}{2} \\ &= P_c \frac{D^4}{8C^4} \cdot \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{\omega_n^2} \cdot a_m^2 f(\omega_n) \\ &= P_c \frac{D^4}{8C^4} \cdot \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{(\omega_s - \omega_m)^2} a_m^2 f(\omega_s - \omega_m). \quad (7.2) \end{aligned}$$

By giving different values to ω_m , different sideband groups, each covering the range ω_s to $\omega_s + 1$, are obtained. The total mean-squared value of all these groups is equal to the summations of the above expression over the relevant range of ω_m . The upper limit of ω_m is, of course, ω_q . (It may be noted that if $\omega_s < \omega_1 + \omega_q$, $f(\omega_s - \omega_m)$ will be zero until ω_m reaches the value $\omega_s - \omega_1$.) The summation is carried out down to $\omega_m = \omega_s/2$, for reasons which will appear in the next paragraph.

Hence the total mean-squared value of the sidebands per unit of bandwidth is

$$\begin{aligned} & P_c \frac{D^4}{8C^4} \sum_{\omega_s/2}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{(\omega_s - \omega_m)^2} a_m^2 f(\omega_s - \omega_m) \\ &= P_c \frac{D^4}{4C^4} \int_{\omega_s/2}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{(\omega_s - \omega_m)^2} \cdot f(\omega_m) \cdot f(\omega_s - \omega_m) d\omega_m, \quad (7.3) \end{aligned}$$

which is a function of ω_s , but not of ω_m .

The summation is stopped at $\omega_s/2$, because extending the summation beyond this point results in a repetition of a corresponding part of the summation in the range $\omega_s/2$ to ω_q . An example may make this clear.

The value of expression (7.2) in the range

$$\omega_m = \frac{\omega_s}{2} - \omega_r \text{ to } \omega_m = \frac{\omega_s}{2} - \omega_r + 1$$

is
$$P_c \frac{D^4}{4C^4} \int_{\frac{\omega_s}{2} - \omega_r}^{\frac{\omega_s}{2} - \omega_r + 1} \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{(\omega_s - \omega_m)^2} \cdot f(\omega_m) \cdot f(\omega_s - \omega_m) d\omega_m.$$

Writing $\omega_n = \omega_s - \omega_m$ this becomes

$$P_c \frac{D^4}{4C^4} \int_{\frac{\omega_s}{2} + \omega_r - 1}^{\frac{\omega_s}{2} + \omega_r} \frac{\omega_q^2}{\omega_n^2} \cdot \frac{\omega_q^2}{(\omega_s - \omega_n)^2} \cdot f(\omega_n) \cdot f(\omega_s - \omega_n) d\omega_n,$$

which is the value of expression (7.2) in the range $\frac{\omega_s}{2} + \omega_r - 1$ to $\frac{\omega_s}{2} + \omega_r$.

Sidebands of frequency $\omega_c + \omega_m - \omega_n$ will now be considered. These sidebands cover the frequency range ω_c to $\omega_c + \omega_q - \omega_1$. The algebra is exactly the same as in the case just considered, except that ω_m is always greater than ω_s , and the limits of ω_m are in this case $\omega_1 + \omega_s$ and ω_q . Corresponding to expression (7.3) for the sum frequency spectrum, the expression for the difference frequency spectrum is

$$P_c \frac{D^4}{4C^4} \int_{\omega_1 + \omega_s}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{(\omega_m - \omega_s)^2} \cdot f(\omega_m) f(\omega_m - \omega_s) d\omega_m. \quad (7.4)$$

The complete expression for the second-order sideband spectrum is the sum of expressions (7.3) and (7.4). Equating this sum to the second term of (4.6),

$$g_2(D, \omega_s) = P_c \frac{D^4}{4C^4} F_2(\omega_s), \quad (7.5)$$

where
$$F_2(\omega_s) = \int_{\omega_s/2}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{(\omega_s - \omega_m)^2} f(\omega_m) f(\omega_s - \omega_m) d\omega_m \\ + \int_{\omega_1 + \omega_s}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{(\omega_m - \omega_s)^2} f(\omega_m) f(\omega_m - \omega_s) d\omega_m. \quad (7.6)$$

The total mean-squared value of the sidebands in the second order spectrum is the sum of the mean-squared values of the sidebands in the two sub-spectra represented by expressions (7.3) and (7.4).

Taking the sidebands of frequency less than ω_c into account, the total mean-squared value of the sidebands in the sub-spectrum represented by (7.3) is

$$P_c \frac{D^4}{2C^4} \int_{2\omega_1}^{2\omega_q} d\omega_s \int_{\omega_s/2}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{(\omega_s - \omega_m)^2} f(\omega_m) f(\omega_s - \omega_m) d\omega_m.$$

Changing the order of integration and writing $\omega_p = \omega_s - \omega_m$, the value of this integral is found to be

$$P_c \frac{D^4}{2C^4} \int_{\omega_1}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} f(\omega_m) d\omega_m \int_{\omega_1}^{\omega_m} \frac{\omega_q^2}{\omega_p^2} f(\omega_p) d\omega_p. \quad (7.7)$$

Similarly, the total mean-squared value of the sidebands in the second sub-spectrum represented by expression (7.4) is

$$P_c \frac{D^4}{2C^4} \int_0^{\omega_q - \omega_1} d\omega_s \int_{\omega_1 + \omega_s}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} \cdot \frac{\omega_q^2}{(\omega_m - \omega_s)^2} \cdot f(\omega_m) f(\omega_m - \omega_s) d\omega_m.$$

Changing the order of integration and writing $\omega_p = \omega_m - \omega_s$, this becomes

$$P_c \frac{D^4}{2C^4} \int_{\omega_1}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} f(\omega_m) d\omega_m \int_{\omega_1}^{\omega_m} \frac{\omega_q^2}{\omega_p^2} f(\omega_p) d\omega_p. \quad (7.8)$$

Hence, the total mean-squared value of the sidebands in the second-order spectrum is the sum of expressions (7.7) and (7.8), and is

$$\begin{aligned} & P_c \frac{D^4}{C^4} \int_{\omega_1}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} f(\omega_m) d\omega_m \int_{\omega_1}^{\omega_m} \frac{\omega_q^2}{\omega_p^2} f(\omega_p) d\omega_p \\ &= P_c \frac{D^4}{C^4} \int_{\omega_1}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} f(\omega_m) d\omega_m \int_{\omega_1}^{\omega_m} \frac{\omega_q^2}{\omega_m^2} f(\omega_m) d\omega_m. \end{aligned}$$

This integral is a particular case of a more general form of multiple integral⁽²⁾. It can be shown that the integral is equal to

$$\begin{aligned} & P_c \frac{D^4}{2C^4} \left[\int_{\omega_1}^{\omega_q} \frac{\omega_q^2}{\omega_m^2} f(\omega_m) d\omega_m \right]^2 \\ &= P_c \frac{K^2 D^4}{2C^4} \quad (8.1) \end{aligned}$$

This is the second term in expression (5.3).

8. Higher Order Spectra.

It is obvious that the distribution of energy in the higher order side-band spectra might be calculated by an extension of the method used in the preceding section. Each spectrum would be divided into a number of sub-spectra corresponding to all the combinations of the positive and negative values of α_m . The calculations, however, become very tedious, and the evaluation of practical cases even more tedious. The first- and second-order spectra which have been calculated are sufficient to allow the complete spectrum to be obtained with reasonable accuracy in certain practical cases, one of which will now be considered.

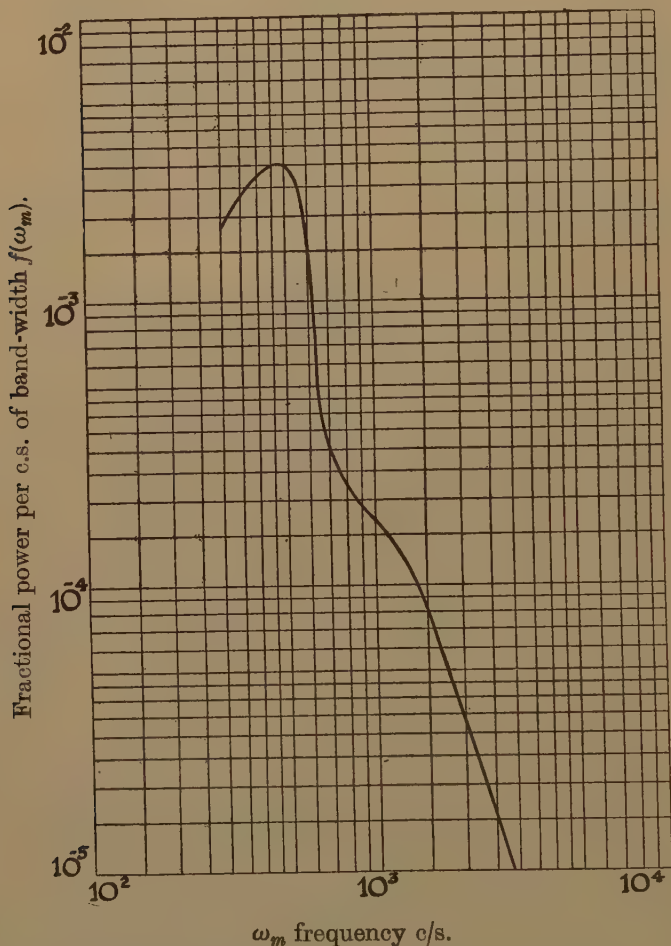
9. Frequency Modulation by Telephonic Signals.

In telephonic communication the signal to be transmitted is conversational speech. Signals of this type may be regarded as the sum

of a very large number of sinusoidal components of unrelated frequencies having amplitudes which are small compared with the amplitude of the whole wave*.

The frequency spectrum of conversational speech extends over a wide range, but in practice, for various reasons, the frequency range is usually limited approximately to the range 300–3000 c/s, this range being sufficient for reasonably high communication efficiency. In practice

Fig. 1.



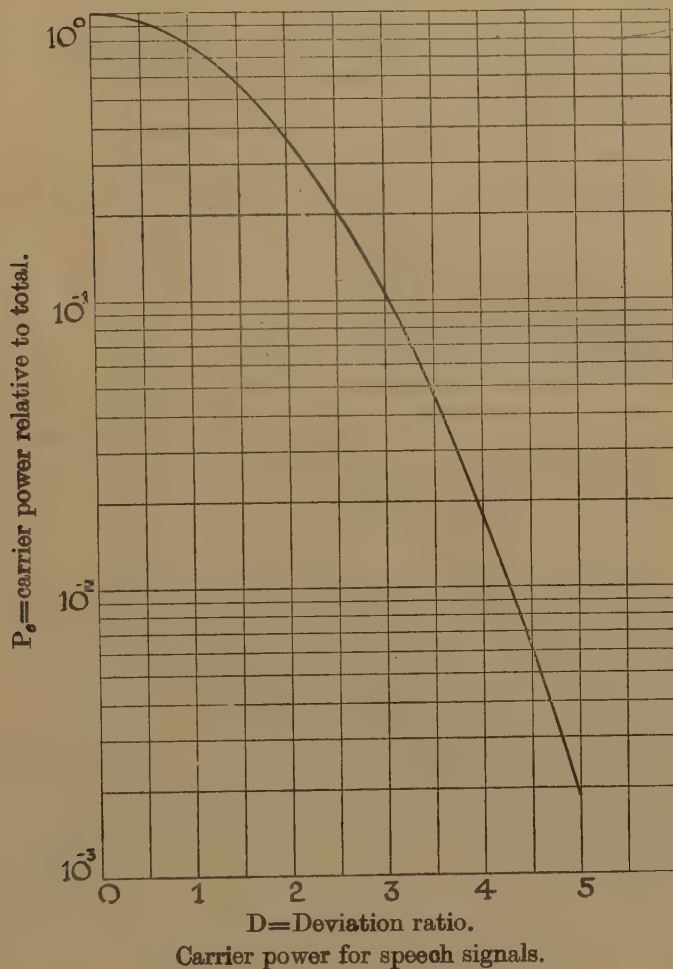
Energy distribution in conversational speech.

* The validity of this assumption will be examined more closely in Part II. of this paper. Actual speech waves differ from the modulating wave described in Section 3, mainly in their distribution of amplitude with time. This implies that in a speech wave the frequencies of the various components are not entirely unrelated. The modifications to the elementary theory given in Part I., which are required to account for these differences, will be considered.

the frequency range does not terminate abruptly at 300 and 3000 c/s, but for the purpose of computation it will be supposed that the frequencies in the speech wave lying between 300 and 3000 c/s are transmitted uniformly, and that all frequencies lying beyond this range are suppressed.

The distribution of energy throughout the entire frequency range of conversational speech has been investigated by Dunn and White⁽³⁾,

Fig. 2.



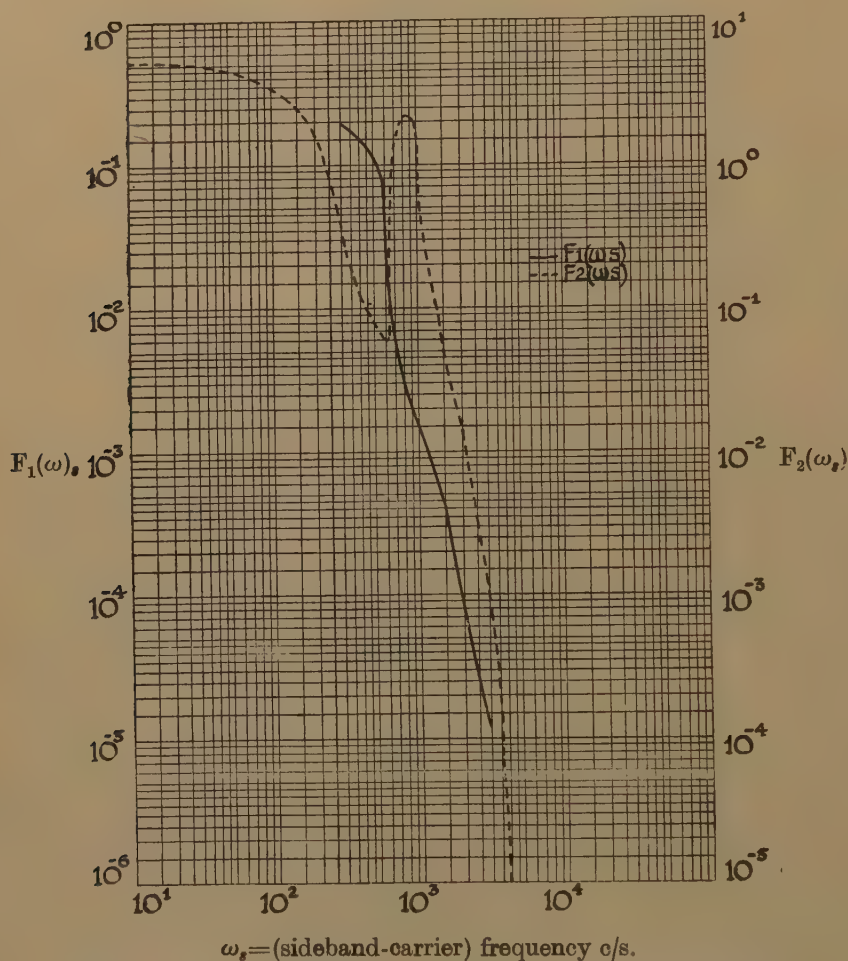
from whose results the graph of fig. 1 has been constructed. This shows the average energy distribution for six male speakers in the frequency band 300–3000 c/s, the ordinate being the mean-squared value per cycle per second of band-width relative to the total mean-squared value.

The quantity
$$K = \int_{\omega_1}^{\omega_2} \frac{\omega^2}{\omega_m^2} f(\omega_m) d\omega_m$$

can be evaluated from fig. 1, and is equal to 40.

Dunn and White⁽³⁾ have also investigated the distribution of amplitude with time in conversational speech. Of the peak values which occur in one-eighth second intervals (the average duration of a syllable), the value 22 DB above the R.M.S. value (crest factor=12.6) is exceeded in only 1 per cent. of all the intervals. Taking this value as the practical upper limit of the crest factor, the mean-squared value of the carrier

Fig. 3.



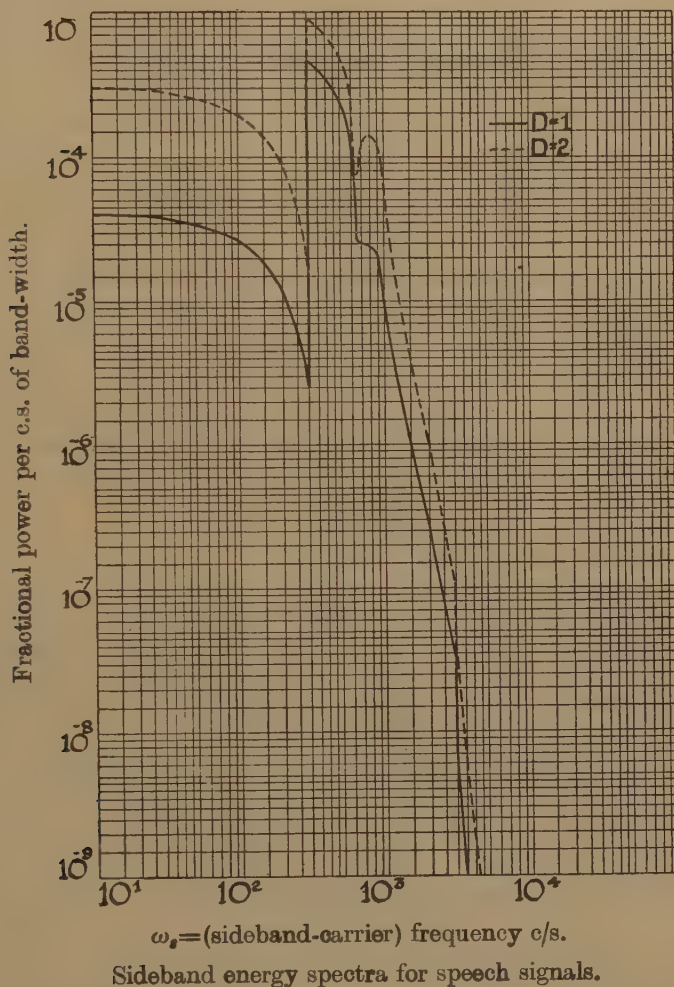
Energy distribution functions for first- and second-order spectra
(speech signals).

relative to the whole wave is (from 5.1) $\exp(-0.25 D^2)$. The graph of this function is shown in fig. 2.

The functions $F_1(\omega_s)$ and $F_2(\omega_s)$ given by expressions (6.6) and (7.6) are shown in fig. 3. Both these functions are derived from fig. 1, the latter by graphical integration.

In order to estimate the accuracy obtainable by using only $F_1(\omega_s)$ and $F_2(\omega_s)$ for the calculation of the complete sideband spectrum, the sum of the mean-squared value of the carrier wave, P_c , and the first two terms of expression (5.3) corresponding to the sum of the powers in the carrier and in the first and second sideband spectra are calculated. This sum

Fig. 4.



for $D=1$ is 0.998 and for $D=2$ is 0.93. For greater values of D the error increases rapidly. $F_1(\omega_s)$ and $F_2(\omega_s)$ having been found, the complete sideband spectrum can easily be calculated from expressions (6.5) and (7.5). Spectra for $D=1$ and $D=2$ are shown in fig. 4. The contribution of the second-order spectrum, even in the case of $D=1$, is considerable. This spectrum does not, of course, stop at the limits shown in fig. 4, but extends from 0 to 6000 c/s.

It should be pointed out that the value of 12.6, which has been taken for the crest factor, is greater than the figure usually employed in design calculations. Values of six or even less are commonly assigned to the crest factor. The reason is that the momentary distortion caused by limiting a few peaks is hardly noticeable in telephone communication. On this new figure, the spectrum of fig. 4 corresponding to $D=2$ would correspond instead to $D=0.95$.

It is interesting to compare the relative sideband energy distribution for a frequency modulated wave with that for an amplitude modulated wave of the same unmodulated carrier power. For the amplitude modulated wave the sideband energy distribution curve is obviously the same as fig. 1, with the ordinate scale multiplied by a constant depending on the degree of modulation. Taking the crest factor of the speech wave as 6, this factor is $\frac{1}{2} \cdot \frac{1}{6^2} = 0.014$. The energy density at $\omega_s = 3000$ c/s is then 1.4×10^{-7} and at $\omega_s = 500$ c/s is 4.2×10^{-5} , both values being relative to the unmodulated carrier power. The corresponding figures for the equivalent case of frequency modulation ($D=2$ in fig. 4) are 10^{-7} at $\omega_s = 3000$ c/s and 9×10^{-4} at $\omega_s = 300$ c/s, the maximum being at 300 c/s in this case.

In the amplitude modulated wave the sideband spectrum ends at $\omega_s = 3000$ c/s, whereas in the frequency modulated wave there is a small spread of energy at higher frequencies.

References.

- (1) E. C. Cherry and R. S. Rivlin, "Non-linear distortion with particular reference to the theory of frequency modulated waves, Part I.," *Phil. Mag.* xxxii. p. 265, October 1941.
- (2) A. S. Gladwin, "A note on a certain multiple integral." *Phil. Mag.* xxxv. p. 657, October 1944.
- (3) H. K. Dunn and S. D. White, "Statistical measurements on conversational speech," *J. Acoust. Soc. Amer.* xi. p. 278, Jan. 1940.

XCIII. Diffusion in Spherical Shells, and a New Method of Measuring the Thermal Diffusivity Constant.

By R. M. BARRER, D.Sc., Ph.D., F.R.I.C.*

[Received June 14, 1944.]

MANY problems of diffusion of matter or heat are directly important in physical and biological sciences. It is the purpose of this paper to provide a group of exact solutions of the diffusion equation, of practical importance and applicable to spherical shells and spheres. A further

* Communicated by the Author.

purpose is to outline a new method of considerable simplicity for determining the thermal diffusivity constant, based upon certain of these equations.

Examples of processes in which diffusion may be the limiting factor are: the permeability of bubble films, and the diffusion of gas or heat through and into spherical shells; solvent extraction by one liquid of a solute dissolved in a second liquid which is immiscible in the first and is dispersed in it as an emulsion; hydrolysis in water of an ester dissolved in a hydrophobic medium dispersed as droplets in the water; nitriding or carburizing spherical iron shells, or spheres such as ball-bearings. These illustrations indicate several of the directions in which solutions of the radial diffusion equation are required. Other examples will be given in the paper, with appropriate solutions.

The General Diffusion Problem and its Solution.

We may restrict ourselves to diffusion in isotropic media. Since diffusion in the sphere is a special case of diffusion in the spherical shell, all solutions in particular problems for the sphere will follow from those for the shell. The general solution for the shell will therefore be given.

Suppose that initially concentrations (or temperatures) c_1 and c_2 are maintained at the surfaces of a spherical shell of inner and outer radii b and a . Then the problem may be formulated as

$$\left. \begin{aligned} \frac{\partial c}{\partial t} &= D \left(\frac{\partial^2 c}{\partial r^2} + \frac{2}{r} \cdot \frac{\partial c}{\partial r} \right), \\ c &= c_1 \text{ at } r = b \text{ for all } t, \\ c &= c_2 \text{ at } r = a \text{ for all } t, \\ c &= f(r) \text{ at } t = 0 \text{ for } b < r < a. \end{aligned} \right\} \dots \dots \dots (1)$$

This may be transformed into a problem involving flow in a plate of thickness $a-b$ (the "equivalent plate") under specific boundary conditions, by making the substitution $u=cr$. Then (1) becomes

$$\left. \begin{aligned} \frac{\partial u}{\partial t} &= D \cdot \frac{\partial^2 u}{\partial t^2}, \\ u_1 &= c_1 b \text{ at } r = b \text{ for all } t, \\ u_2 &= c_2 a \text{ at } r = a \text{ for all } t, \\ u &= rf(r) \text{ for } b < r < a \text{ at } t = 0. \end{aligned} \right\} \dots \dots \dots (2)$$

For this problem the general solution is known to be ⁽¹⁾

$$\begin{aligned} u &= u_1 + \frac{(u_2 - u_1)}{(a-b)}(r-b) + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{u_2 \cos n\pi - u_1}{n} \cdot \sin \frac{n\pi(r-b)}{(a-b)} \exp \left(-\frac{Dn^2\pi^2 t}{(a-b)^2} \right) \\ &+ \frac{2}{(a-b)} \sum_{n=1}^{\infty} \sin \frac{n\pi(r-b)}{(a-b)} \exp \left(-\frac{Dn^2\pi^2 t}{(a-b)^2} \right) \int_{r=b}^{r'=a} r' f(r') \sin \frac{n\pi(r'-b)}{(a-b)} dr'. \end{aligned} \dots \dots \dots (3)$$

Back substitution then leads to the general solution for the spherical shell :

$$c = \frac{c_1 b}{r} + \frac{c_1 a - c_1 b}{(a-b)} \cdot \frac{(r-b)}{r} + \frac{2}{\pi r} \sum_{n=1}^{\infty} \frac{c_2 a \cos n\pi - c_1 b}{n} \sin \frac{n\pi(r-b)}{(a-b)} \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right) + \frac{2}{r(a-b)} \sum_{n=1}^{\infty} \sin \frac{n\pi(r-b)}{(a-b)} \cdot \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right) \int_{r'=b}^{r'=a} r' f(r') \sin \frac{n\pi(r'-b)}{(a-b)} \cdot dr'. \quad (4)$$

Particular Problems concerning Spherical Shells.

As a rule the initial concentration within the shell is constant : $f(r)=c_0$. Equation (4) then becomes

$$c = \frac{c_1 b}{r} + \frac{c_2 a - c_1 b}{(a-b)} \cdot \frac{(r-b)}{r} + \frac{2}{\pi r} \sum_{n=1}^{\infty} \frac{c_2 a \cos n\pi - c_1 b}{n} \cdot \sin \frac{n\pi(r-b)}{(a-b)} \cdot \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right) - \frac{2c_0}{\pi r} \sum_{n=1}^{\infty} \sin \frac{n\pi(r-b)}{(a-b)} \cdot \frac{(a \cos n\pi - b)}{n} \cdot \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right). \quad (5)$$

This solution will apply for example to the uptake of gas * by spherical metallic shells with a given initial gas content (concentration c_0), and with constant gas concentrations at the two surfaces (c_1 at $r=b$; c_2 at $r=a$).

It is possible in other instances that the concentration of gas initially within the shell may not be independent of r . Thus when $c=c_0/r$ for $b < r < a$ at $t=0$, the appropriate solution is

$$c = \frac{c_1 b}{r} + \frac{c_2 a - c_1 b}{(a-b)} \cdot \frac{(r-b)}{r} + \frac{2}{\pi r} \sum_{n=1}^{\infty} \frac{c_2 a \cos n\pi - c_1 b}{n} \cdot \sin \frac{n\pi(r-b)}{(a-b)} \cdot \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right) + \frac{4c_0}{\pi r} \sum_{m=0}^{\infty} \sin \frac{(2m+1)\pi(r-b)}{(a-b)} \cdot \frac{1}{(2m+1)} \cdot \exp\left(-\frac{D(2m+1)^2\pi^2 t}{(a-b)^2}\right). \quad (6)$$

Other examples of the application of eq. (4) follow equally directly when $c=f(r)$ for $b < r < a$ and $t=0$ is known, for one may then evaluate

$$\int_{r'=b}^{r'=a} r' f(r') \sin \frac{n\pi(r'-b)}{(a-b)} \cdot dr',$$

in eq. 4 to obtain the solution in forms analogous to (5) or (6).

* In these and all analogous cases the solution assumes that all phase-boundary processes are rapid compared with diffusion. Phase-boundary processes are discussed in ref. 2, chap. 4. It is to be noted that metals are good solvents for some gases. H_2 dissolves in many transition and rare-earth elements⁽²⁾ (Ni, Fe, Cu, Pt, Pd, Ce, Th, V, Ti); O_2 in Ag⁽²⁾, Ti⁽³⁾, Zr⁽⁴⁾; or N_2 in Zr⁽⁴⁾, Mo⁽⁵⁾. The degassing of metals is technically very important⁽⁵⁾.

Particular cases of (5) and (6) are easily derived. When $c_1=c_2=0$ at $r=a$ and b , and when initially for $b < r < a$, $c=c_0$ and c_0/r respectively, one obtains the solutions (5 a) and (6 a) :

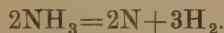
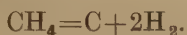
$$c = -\frac{2c_0}{\pi r} \sum_{n=1}^{\infty} \sin \frac{n\pi(r-b)}{(a-b)} \frac{a \cos n\pi - b}{n} \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right), \quad (5a)$$

$$c = \frac{4c_0}{\pi r} \sum_{m=0}^{\infty} \sin \frac{(2m+1)\pi(r-b)}{(a-b)} \cdot \frac{1}{(2m+1)} \exp\left(-\frac{D(2m+1)^2\pi^2 t}{(a-b)^2}\right). \quad (6a)$$

Again when $c_1=c_2$ at $r=a$ and b for all t , and when initially $c=c_0$ for $b < r < a$, then eq. (5) becomes

$$c = c_1 + \frac{2(c_1-c_0)}{\pi r} \sum_{n=1}^{\infty} \frac{a \cos n\pi - b}{n} \sin \frac{n\pi(r-b)}{(a-b)} \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right). \quad (7)$$

Equations (5 a) and (6 a) give the concentration distributions at any time during degassing of metallic shells with initial concentration distributions within the shell wall $c=c_0$ and c_0/r respectively. Equation (7) gives the concentration distribution during carburizing or nitriding of steel shells when the surface concentration of carbon or nitrogen is kept fixed by comparatively rapid surface reactions such as



It equally well describes the distribution of the alloying metal (Zn or Sn) during zinc or tin plating of thin metallic shells by immersion in vapour or liquid baths of the plating metals. Usually in this case the initial concentration c_0 of tin or zinc within the shell is zero. Complications may, however, arise when the alloying metal forms intermetallic compounds with the solvent metal.

The equations (5 a) and (7) also serve to give temperature distributions during cooling or heating of spherical shells in the absence of emissivity effects at the surface. Such effects will normally be absent when the surfaces at $r=a$ and b are bounded by liquids at constant temperature and in thermal equilibrium with the interfaces.

The Sphere.

Corresponding to the general solution (4) the general solution for the sphere is obtained by setting $b=0$, and $c_1=0$:

$$c = c_2 + \frac{2}{\pi r} \sum_{n=1}^{\infty} \frac{c_2 a \cos n\pi}{n} \cdot \sin \frac{n\pi r}{a} \cdot \exp\left(-\frac{Dn^2\pi^2 t}{a^2}\right) \\ + \frac{2}{\pi a} \sum_{n=1}^{\infty} \sin \frac{n\pi r}{a} \cdot \exp\left(-\frac{Dn^2\pi^2 t}{a^2}\right) \int_{r'=0}^{r'=a} r' f(r') \sin \frac{n\pi r'}{a} dr'. \quad (4a)$$

Equations for diffusion in the sphere for the particular cases dealt with in the previous section then follow from (4 a) in the same way as the solutions in that section were derived from (4). Thus the diffusion of carbon into steel bearings containing initially a uniform carbon

concentration c_0 , and with a constant concentration c_2 of carbon at the surface, follows the equation :

$$c = c_2 + \frac{2(c_2 - c_0)}{\pi r} \sum_{n=1}^{\infty} \frac{a \cos n\pi}{n} \cdot \sin \frac{n\pi r}{a} \cdot \exp\left(-\frac{Dn^2\pi^2 t}{a^2}\right). \quad (7a)$$

Steady State Distribution of Concentration or Temperature across the Shell.

This distribution also follows easily from the general solution (4), since this state is established when $t \rightarrow \infty$, giving the following distribution which is independent of time :

$$c = \frac{c_1 b}{r} + \frac{c_2 a - c_1 b}{(a-b)} \cdot \frac{(r-b)}{r}, \quad \dots \dots \dots (8)$$

or

$$\frac{c_1 - c}{c_1 - c_2} = \frac{a(r-b)}{r(a-b)} \quad \dots \dots \dots (8a)$$

Rate of Flow across the Shell.

From eq. (5) the concentration gradient at any point r is found, by differentiation, to be

$$\begin{aligned} \frac{dc}{dr} = & -\frac{c_1 b}{r^2} + \frac{(c_2 a - c_1 b)}{(a-b)} \frac{b}{r^2} \\ & + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{c_2 a \cos n\pi - c_1 b}{n} \left[\frac{n\pi}{(a-b)} \cdot \frac{1}{r} \cdot \cos \frac{n\pi(a-b)}{(a-b)} \right. \\ & \quad \left. - \frac{1}{r^2} \sin \frac{n\pi(r-b)}{a-b} \right] \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right) \\ & - \frac{2c_0}{\pi} \sum_{n=1}^{\infty} \frac{a \cos n\pi - b}{n} \left[\frac{n\pi}{(a-b)} \cdot \frac{1}{r} \cdot \cos \frac{n\pi(a-b)}{(a-b)} \right. \\ & \quad \left. - \frac{1}{r^2} \sin \frac{n\pi(r-b)}{(a-b)} \right] \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right). \quad (9) \end{aligned}$$

At the two boundaries

$$\begin{aligned} \left(\frac{dc}{dr}\right)_{r=a} = & -\frac{c_1 b}{a^2} + \frac{c_2 a - c_1 b}{(a-b)} \cdot \frac{b}{a^2} \\ & + \frac{2}{(a-b)} \sum_{n=1}^{\infty} \frac{c_2 a \cos n\pi - c_1 b}{a} \cdot \cos n\pi \cdot \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right) \\ & - \frac{2c_0}{(a-b)} \sum_{n=1}^{\infty} \frac{a \cos n\pi - b}{a} \cdot \cos n\pi \cdot \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right), \quad \dots \dots (9a) \end{aligned}$$

$$\begin{aligned} \left(\frac{dc}{dr}\right)_{r=b} = & -\frac{c_1}{b} + \frac{c_2 a - c_1 b}{(a-b)} \cdot \frac{1}{b} \\ & + \frac{2}{(a-b)} \sum_{n=1}^{\infty} \frac{c_2 a \cos n\pi - c_1 b}{b} \cdot \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right) \\ & - \frac{2c_0}{(a-b)} \sum_{n=1}^{\infty} \frac{a \cos n\pi - b}{b} \exp\left(-\frac{Dn^2\pi^2 t}{(a-b)^2}\right). \quad \dots \dots (9b) \end{aligned}$$

The rate of flow across the boundary surfaces is then

$$q_a = -4\pi a^2 D \left(\frac{dc}{dr} \right)_{r=a}, \quad \dots \dots \dots (10a)$$

and
$$q_b = -4\pi b^2 D \left(\frac{dc}{dr} \right)_{r=b} \dots \dots \dots (10b)$$

at $r=a$ and $r=b$ respectively. Substitution of the differentials (9 a) and (9 b) in (10 a) and (10 b) then gives the flux across the two boundaries. In the steady state (i. e. as $t \rightarrow \infty$), this flux is

$$q_s = \frac{4\pi D \cdot ab}{(a-b)} (c_2 - c_1) \dots \dots \dots (10c)$$

Quantity of Solute or Heat absorbed in the Shell.

The total amount of solute which has flowed into or out of the shell in time t is

$$Q_1 = -4\pi D \cdot \int_0^t \left[b^2 \left(\frac{dc}{dr} \right)_{r=b} - a^2 \left(\frac{dc}{dr} \right)_{r=a} \right] dt. \quad \dots \dots (11)$$

Substitution of (9 a) and (9 b) into (11), and integration, then gives

$$\begin{aligned} Q_1 = & \frac{4}{3} \pi (a-b) [a^2 c_2 - (a^2 + b^2) c_0 + b^2 c_1] \left[1 - \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp \left(-\frac{D n^2 \pi^2 t}{(a-b)^2} \right) \right] \\ & + \frac{4}{3} \pi ab(a-b) \left[\frac{c_2}{2} - c_0 + \frac{c_1}{2} \right] \left[1 - \frac{12}{\pi^2} \sum_{n=1}^{\infty} \frac{\cos n\pi}{n^2} \exp \left(-\frac{D n^2 \pi^2 t}{(a-b)^2} \right) \right]. \end{aligned} \quad \dots \dots (12)$$

The total amount of solute remaining in the shell after this time interval is

$$Q_2 = \frac{4\pi}{3} (a^3 - b^3) c_0 + Q_1, \quad \dots \dots \dots (13)$$

where Q_1 is given by eq. (12). When c_1 and c_2 are both zero, eq. (12) reduces to

$$\begin{aligned} Q_3 = & \frac{4\pi}{3} (a^3 - b^3) c_0 - \frac{4\pi}{3} (a-b)(a^2 + b^2) c_0 \left[1 - \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp \left(-\frac{D n^2 \pi^2 t}{(a-b)^2} \right) \right] \\ & - \frac{4\pi}{3} ab(a-b) c_0 \left[1 - \frac{12}{\pi^2} \sum_{n=1}^{\infty} \frac{\cos n\pi}{n^2} \exp \left(-\frac{D n^2 \pi^2 t}{(a-b)^2} \right) \right], \end{aligned} \quad \dots \dots (14)$$

so that the amount of solute desorbed is then

$$Q_4 = \frac{4\pi}{3} (a^3 - b^3) c_0 - Q_3. \quad \dots \dots \dots (15)$$

Finally, when $c_0=0$ and $c_1=c_2$, eq. (12) becomes

$$\begin{aligned} Q_5 = & \frac{4\pi}{3} (a-b)(a^2 + b^2) c_1 \left[1 - \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp \left(-\frac{D n^2 \pi^2 t}{(a-b)^2} \right) \right] \\ & + \frac{4\pi}{3} ab(a-b) c_1 \left[1 - \frac{12}{\pi^2} \sum_{n=1}^{\infty} \frac{\cos n\pi}{n^2} \exp \left(-\frac{D n^2 \pi^2 t}{(a-b)^2} \right) \right]. \end{aligned} \quad \dots \dots (16)$$

Such equations are very useful in determining the amounts of gas removed from or remaining in metallic shells during degassing. They may also be used in quantitative interpretations of the drying of shells of rubber or plastics impregnated by water or organic solvents, and of the uptake of such liquids by the shells. It is necessary for quantitative applications that the flow of solute should be governed only by diffusion, and that D should not depend on the concentration of the solute.

The equations are equally useful in obtaining quantities of heat in shells during heating or cooling in the absence of emissivity effects, for example in contact with stirred liquids at constant temperature. In

these applications, however, Q must be replaced by $\frac{Q}{c_p \rho}$ *, where c_p is the specific heat and ρ the density of the medium; while the concentration c must be replaced by the temperature T .

The Quantity of Solute or Heat in the Sphere.

The quantity of a solute in a sphere for the conditions given in the previous section can be obtained quite simply. To derive solutions corresponding to equations (12) to (16) it is only necessary to set $c_1=0$ and $b=0$ in each. For example, the quantity of solute in a sphere after a time t when $c=c_2$ at $r=a$, and when $c=c_0$ at $t=0$ for $0 < r < a$, is from eq. (13)

$$Q_6 = \frac{4\pi}{3} a^3 c_0 + \frac{4\pi}{3} a^3 (c_2 - c_0) \left[1 - \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp\left(-\frac{Dn^2\pi^2 t}{a^2}\right) \right]. \quad (13a)$$

When $c_2=0$ at $r=a$ for all t , and $c=c_0$ at $t=0$ between $0 < r < a$, one obtains for Q_7 the amount of solute remaining in the sphere after an interval t ,

$$Q_7 = \frac{4\pi}{3} \cdot a^3 c_0 \left[\frac{6}{\pi^2} \cdot \sum_{n=1}^{\infty} \frac{1}{n^2} \cdot \exp\left(-\frac{Dn^2\pi^2 t}{a^2}\right) \right]. \quad (14a)$$

Other solutions may be derived for appropriate conditions equally readily.

Time Lag in establishing the Steady Permeation Rate.

When heat or matter diffuses through a spherical shell the flux on the outgoing side of the shell varies with time and finally approaches a steady value. If the total flow Q is plotted against time (fig. 1) a straight line is therefore eventually obtained. When this asymptote is extrapolated backward it makes intercepts L and L_1 on the axes of t and Q respectively, which may be calculated, using equations already derived. Thus at the

* This is because for a volume V of the medium one has for a solute

$$Q = c \cdot V$$

while in the thermal case

$$Q = c_p \cdot \rho \cdot V \cdot T.$$

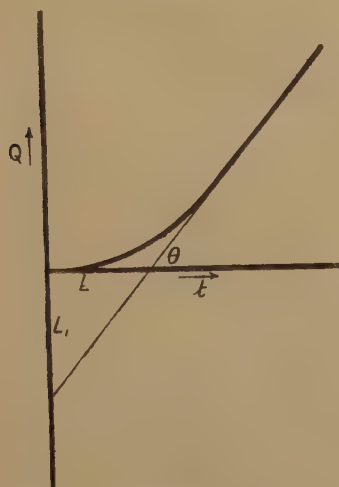
outgoing surface $r=a$ the flux q_a of solute is given by eq. (9 a). The total flow across this surface is therefore

$$Q_s = \int_0^t q_a \cdot dt, \quad \dots \dots \dots (17)$$

so that substituting for q_a in eq. (17) and integrating one finds

$$\begin{aligned} Q_s = 4\pi \cdot D \cdot (c_1 - c_2) \frac{ab}{(a-b)} t \\ - \frac{4\pi}{3} a^2(a-b)(c_2 - c_0) \left[1 - \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp \left(- \frac{Dn^2\pi^2 t}{(a-b)^2} \right) \right] \\ - \frac{2\pi}{3} ab(a-b)(c_1 - c_0) \left[1 - \frac{12}{\pi^2} \sum_{n=1}^{\infty} - \frac{\cos n\pi}{n^2} \exp \left(- \frac{Dn^2\pi^2 t}{(a-b)^2} \right) \right]. \end{aligned} \quad (18)$$

Fig. 1.



As t becomes large the curve showing Q vs. t (fig. 1) thus approaches the asymptote

$$Q_s = 4\pi D \frac{ab}{(a-b)} (c_1 - c_2) t - \frac{4\pi}{3} a^2(a-b)(c_2 - c_0) - \frac{2\pi}{3} ab(a-b)(c_1 - c_0), \quad (19)$$

$$\text{i. e. } Q_s = 4\pi D \frac{ab}{(a-b)} (c_1 - c_2) \left[t - \frac{c_1 - c_0}{c_1 - c_2} \cdot \frac{(a-b)^2}{6D} - \frac{c_2 - c_0}{c_1 - c_2} \cdot \frac{a}{b} \cdot \frac{(a-b)^2}{3D} \right]. \quad (19a)$$

Thus the intercept L on the time axis is

$$L = \frac{c_1 - c_0}{c_1 - c_2} \cdot \frac{(a-b)^2}{6D} + \frac{c_2 - c_0}{c_1 - c_2} \cdot \frac{a}{b} \cdot \frac{(a-b)^2}{3D}. \quad \dots \dots \dots (20)$$

When $c_2 = c_0$ one obtains the very simple result

$$L = \frac{(a-b)^2}{6D}. \quad \dots \dots \dots (20a)$$

Finally, the intercept L_1 on the axis of Q is given by $L_1 = -L \tan \theta$.

Since

$$\tan \theta = 4\pi D \frac{ab}{(a-b)} (c_1 - c_2)$$

(eq. 10 c) one obtains

$$L_1 = -\frac{2\pi}{3} \cdot ab(a-b)(c_1 - c_0) - \frac{4\pi}{3} \cdot a^2(a-b)(c_2 - c_0). \quad (21)$$

A Method for Measuring the Thermal Diffusivity Constant.

The author has previously calculated the intercept L of fig. 1 for diffusion through the plate ⁽⁶⁾ and hollow tube ⁽⁷⁾, and used the value of L to obtain the diffusion constant of solutes in solid media. This paper now gives the intercept for the spherical shell, and it is clear that the measurement of L may also give the thermal diffusivity constant (K). Replacing D by K , and concentration (c) by temperature (T), one has for the plate of thickness l , the spherical shell and the hollow tube respectively :

$$\left. \begin{aligned} L &= \frac{T_1 - T_0}{T_1 - T_2} \cdot \frac{l^2}{6K} + \frac{T_2 - T_0}{T_1 - T_2} \cdot \frac{l^2}{3K} \\ L &= \frac{T_1 - T_0}{T_1 - T_2} \cdot \frac{(a-b)^2}{6K} + \frac{T_2 - T_0}{T_1 - T_2} \cdot \frac{a}{b} \cdot \frac{(a-b)^2}{3K} \\ L &= -\frac{\pi^2 a}{2K} \frac{\ln b/a}{T_1 - T_2} \sum_{n=1}^{\infty} \frac{J_0^2(\alpha_n a)}{J_0^2(\alpha_n b) - J_0^2(\alpha_n a)} [U_0'(\alpha_n r)]_{r=a} \cdot I, \end{aligned} \right\} \quad (22)$$

where a and b are the external and internal radii of tube or shell,

$$I = \int_b^a r \cdot U_0(\alpha_n r) \left[T_0 - \frac{(T_1 - T_2) \ln r + T_2 \ln b - T_1 \ln a}{\ln b - \ln a} \right] dr.$$

$$U_0(\alpha r) = J_0(\alpha r) H_0'(\alpha b) - J_0(\alpha b) H_0'(\alpha r).$$

$$U_0'(\alpha r) = \frac{d}{dr} U_0(\alpha r).$$

α_n = the n th positive root of the equation $U_0(\alpha_n b) = 0$.

In the definition of $U_0(\alpha r)$, J_0 and H_0' denote Bessel functions of the first and the third kind, and of zero order. The root α_n may be determined from the formula given by Gray, Mathews and MacRobert ⁽⁸⁾, and the Bessel functions evaluated from tables. Arithmetical manipulation of the equation for hollow tubes is, however, less simple than the use of the equations for plates and shells, and we will in what follows consider only the latter.

These equations take a very simple form when $T_2 = T_0 = T$, which may very conveniently be room temperature. One then has

$$L = \frac{l^2}{6K} \quad \dots \dots \dots (23)$$

$$\text{(for the plate),} \quad L = \frac{(a-b)^2}{6K} \quad \dots \dots \dots (20a)$$

(for the spherical shell).

The intercept L does not now depend on T_1 , T_2 or T_0 . The same value of L is obtained, moreover, when any property of the system proportional to Q is plotted against t (fig. 1). For instance one may have stirred liquids at temperatures T_1 and T_2 on ingoing and outgoing sides of the plate or shell. One may then measure the rise in temperature of the fluid at T_2 provided that this rise in T_2 during the establishing of the steady state is small compared with $(T_1 - T_2)$. Since there is no need to measure the absolute amount of heat transferred the labour both experimental and arithmetical is reduced to a minimum.

The method, which does not appear in standard works⁽⁹⁾, is well suited to measuring K for poor conductors and insulators. The magnitude of L for plates and shells 1 cm. thick is given in typical instances in Table I. L increases as the square of this thickness.

TABLE I.
Intercepts L calculated for Plates or Shells of
1 cm. Thickness of Wall.

Material.	Diffusivity constant (cm. ² /sec.).	L in sec.
Ebonite	0.0010	166.7
Water	0.00143	116.5
Coal	0.002	83.4
Brick	0.005	33.3
Glass	0.0057	29.3
Granite	0.0155	10.8
Mercury	0.033	5.0
Copper	1.133	0.147

Summary.

Solutions of the diffusion equation for flow of heat or matter in spherical shells are derived for a number of cases of practical importance. A new simple procedure is outlined for determining the thermal diffusivity constant, especially of poor conductors and insulators.

References.

- (1) Carslaw, 'Conduction of Heat,' 1921, p. 67. Macmillan and Co.
- (2) Barrer, 'Diffusion in and through Solids,' C.U.P. 1941.
- (3) Ehrlich, *Zeit. anorg. Chem.* cexlvii. p. 53 (1941).
- (4) De Boer and Fast, *Rec. Trav. Chim. Pays-Bas*, lv. p. 459 (1936); *ibid.* lix. p. 161 (1940); Fast, *Metallwirtschaft*, xvii. p. 641 (1938).
- (5) Norton and Marshall, *Metals Technology*, 1944, T.P. 1643.
- (6) Barrer, *Trans. Far. Soc.* xxxv. p. 628 (1939).
- (7) Barrer, *Trans. Far. Soc.* xxxvi. p. 1235 (1940).
- (8) Gray, Mathews and MacRobert, 'Treatise on Bessel Functions.' Macmillan and Co. 1922.
- (9) *e. g.*, Glazebrook, 'Dictionary of Applied Physics.'

The Chemical Laboratories,
The Technical College,
Bradford.

XCIV. *The tabulation of some Bessel Functions $K_\nu(x)$ and $K'_\nu(x)$ of Fractional Order.*

By H. R. F. CARSTEN, Dr.-Ing., F.Inst.P., and
Miss N. W. MCKERROW, B.Sc.*

[Received June 5, 1944.]

I. *Introduction.*

UNDER certain conditions, the temperature field within a cylindrical rod, subjected to a sudden change in temperature, may be developed in terms of modified Bessel functions of the second kind, of order $n \pm \frac{1}{2}$, and their derivatives. As values of such functions do not yet appear to have been published, it has been found desirable to prepare tables of these for a range of the variables.

II. *The Theory in which the Functions appear.*

The two dimensional temperature field $u(r, t)$ in a homogeneous cylindrical rod, of material with the diffusivity α^2 , may be obtained from the heat equation

$$\alpha^2 \Delta u = \frac{\partial u}{\partial t}$$

with

$$\Delta u = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial u}{\partial r} \right),$$

t the time and r the radial distance.

Differential equations of this type are conveniently solved by applying the Laplace transformation to the unknown function

$$\bar{u} = L[u(t)],$$

the operator L implying the transformation ⁽¹⁾

$$L[f(t)] = \int_0^\infty e^{-pt} f(t) dt. \quad R(p) > 0.$$

Under this transformation the original differential equation is replaced by a Bessel type differential equation for the transform

$$\frac{d^2 \bar{u}}{dr^2} + \frac{1}{r} \frac{d\bar{u}}{dr} - \frac{p}{\alpha^2} \bar{u} = 0.$$

If the surface temperature of a cylinder of infinite length is subjected to a sudden change, according to the boundary conditions

$$t=0, \quad u=0, \quad \bar{u}=0, \quad x=\frac{r}{R},$$

* Communicated by J. C. P. Miller.

$$t > 0, x = 1, u = 1, \bar{u} = \frac{1}{p}, R = \text{radius of the cylinder},$$

then the solution is found to be

$$\bar{u} = \frac{1}{p} \frac{J_0\left(x \frac{R}{a} \sqrt{p}\right)}{J_0\left(i \frac{R}{a} \sqrt{p}\right)}; \quad i = \sqrt{-1}.$$

For large values of p , corresponding to small values of t , the denominator can be developed as follows :—

$$\begin{aligned} \frac{1}{p} \frac{1}{J_0\left(i \frac{R}{a} \sqrt{p}\right)} &\sim {}^2 \sqrt{\pi \frac{R}{2a}} p^{-\frac{3}{4}} e^{-\frac{R}{a} \sqrt{p}} \\ &= {}^2 \sqrt{\pi \frac{R}{2a}} \cdot L\left[\frac{t^{-1/4}}{\Gamma(\frac{3}{4})}\right] \cdot L\left[\frac{R}{2a} \frac{e^{-\left(\frac{R}{2a}\right)^2 \frac{1}{t}}}{t \sqrt{\pi t}}\right]. \end{aligned}$$

Applying the Folding Theorem of the Laplace transformation, whereby

$$L[f_1(t)] \cdot L[f_2(t)] = L\left[\int_0^t f_1(\tau) \cdot f_2(t-\tau) d\tau\right],$$

the right hand side becomes

$$\frac{2}{\Gamma(\frac{3}{4})} \left(\frac{R}{2a}\right)^{3/2} \cdot L\left[\int_0^t \frac{e^{-\left(\frac{R}{2a}\right)^2 \frac{1}{t-\tau}}}{(t-\tau)^{3/2} \cdot \tau^{1/4}} d\tau\right],$$

which, on substituting $\tau = t \tanh^2 \frac{\phi}{2}$, $\alpha = 2 \left(\frac{R}{2a}\right)^{-2}$, is equal to

$$\begin{aligned} &\frac{\sqrt{2}}{\Gamma(\frac{3}{4})} \cdot L\left[\frac{e^{-\frac{1}{\alpha t}}}{\left(\frac{\alpha t}{2}\right)^{3/4}} \int_0^\infty e^{-\frac{1}{\alpha t} \cosh \phi} \cdot \sinh^{1/2} \phi d\phi\right] \\ &= {}^2 \sqrt{\frac{2}{\pi}} \cdot L\left[\frac{e^{-\frac{1}{\alpha t}}}{\sqrt{\alpha t}} \left(\frac{1}{2\alpha t}\right)^{1/4} \frac{\sqrt{\pi}}{\Gamma(\frac{3}{4})} \int_0^\infty e^{-\frac{1}{\alpha t} \cosh \phi} \cdot \sinh^{1/2} \phi d\phi\right] \\ &= {}^2 \sqrt{\frac{2}{\pi}} \cdot L\left[\frac{e^{-\frac{1}{\alpha t}}}{\sqrt{\alpha t}} K_{1/4}\left(\frac{1}{\alpha t}\right)\right] \end{aligned}$$

from the integral representation for the modified Bessel function of the second kind ⁽²⁾ :

$$K_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu \sqrt{\pi}}{\Gamma(\nu + \frac{1}{2})} \int_0^\infty e^{-z \cosh \phi} \cdot \sinh^{2\nu} \phi d\phi, \quad \nu = \frac{1}{4}, \quad z = \frac{1}{\alpha t}.$$

For sufficiently small x , the numerator may be expanded into the series

$$J_0\left(ix\frac{R}{a}\sqrt{p}\right)=\sum_{\nu=0}^{\infty}\frac{1}{(\nu!)^2}\left(\frac{2x^2}{\alpha}\right)^{\nu}p^{\nu}.$$

Therefore, for a short period immediately after the impulse, the temperature function for points near the centre of the rod approximates to

$$u(x, t)=2\sqrt{\frac{2}{\pi}}\sum_{\nu=0}^{\infty}\frac{1}{(\nu!)^2}\left(\frac{2x^2}{\alpha}\right)^{\nu}\frac{d^{\nu}}{dt^{\nu}}\left[\frac{1}{\sqrt{\alpha t}}e^{-\frac{1}{\alpha t}}K_{1/4}\left(\frac{1}{\alpha t}\right)\right].$$

$$\alpha=2\left(\frac{R}{2a}\right)^{-2}.$$

For the calculation of $u(t)$ from this expansion, values of the function $K_{n\pm\frac{1}{2}}(x)$ are required. As these do not appear to have been tabulated yet, $K_{\frac{1}{2}}(x)$ and $K_{\frac{3}{2}}(x)$ have been calculated for a range of x , from which $K_{n\pm\frac{1}{2}}(x)$ for higher values of n may be found by means of the recurrence relation

$$K_{\nu-1}(x)-K_{\nu+1}(x)=-\frac{2\nu}{x}K_{\nu}(x).$$

Values of x have been taken from 0 to 5, with interval 0.1, also $x=6, 8, 10$. With a few exceptions the tabulated values are given to five significant figures, the last figure retained being rounded off according to the usual rule. Seven figures were used throughout the computation, and consistency of the results checked by the calculation of difference tables.

III. Procedure of Calculation.

As the functions $K_{n\pm\frac{1}{2}}(x)$ were to be obtained by interpolation with n as variable, the functions $K_{n+\frac{1}{2}}(x)$ were calculated first from the definition

$$K_{n+1/2}(x)=\sqrt{\frac{\pi}{2x}}e^{-x}\sum_{s=0}^n\frac{(n+s)!}{s!(n-s)!}\frac{1}{(2x)^s},$$

whence

$$K_{1/2}(x)=\sqrt{\frac{\pi}{2x}}e^{-x}=\lambda,$$

$$K_{3/2}(x)=\lambda\left(1+\frac{1}{x}\right),$$

$$K_{5/2}(x)=\lambda\left(1+\frac{3}{x}+\frac{3}{x^2}\right), \text{ etc.}$$

Thereafter, $K_{\frac{1}{2}}(x)$, $K_{\frac{3}{2}}(x)$ were computed by Lagrange's interpolation formulæ from tabulated values of K_0 , K_1 , K_2 . . . etc. ⁽³⁾, and previously calculated $K_{\frac{1}{2}}$, $K_{\frac{3}{2}}$. . . etc., using the relation

$$K_{-\nu}(x)=K_{\nu}(x).$$

Consider $K_n(x)$ as a function of n for fixed x , and let n take values indicated in the following table :—

Argument n .	Function $K_n(x)$.
$-\frac{m}{2}$	$K_{-\frac{m}{2}}$
\vdots	\vdots
-1	K_{-1}
$-\frac{1}{2}$	$K_{-\frac{1}{2}}$
0	K_0
$\frac{1}{2}$	$K_{\frac{1}{2}}$
1	K_1
\vdots	\vdots
$\frac{m}{2}$	$K_{\frac{m}{2}}$

then the Lagrange interpolation formula for $(2m+1)$ points is

$$\begin{aligned} K_{n_0}(x) &= \sum_{s=-m}^m \frac{\left(n_0 + \frac{m}{2}\right)\left(n_0 + \frac{m-1}{2}\right) \dots \left(n_0 - \frac{s-1}{2}\right) \times \left(n_0 - \frac{s+1}{2}\right) \dots \left(n_0 - \frac{m}{2}\right)}{\left(\frac{s}{2} + \frac{m}{2}\right)\left(\frac{s}{2} + \frac{m-1}{2}\right) \dots \left(\frac{s}{2} - \frac{s-1}{2}\right) \times \left(\frac{s}{2} - \frac{s+1}{2}\right) \dots \left(\frac{s}{2} - \frac{m}{2}\right)} K_{s/2}(x) \\ &= \sum_{s=-m}^m A_s K_{s/2}(x). \end{aligned}$$

Corresponding positive and negative values of n should be taken, whence the Lagrange coefficients for a certain value of m , $(A_{-m}, A_{-(m-1)}, \dots, A_{-1}, A_0, A_1, \dots, A_m)$ may be combined in pairs to give $(A_0, (A_1 + A_{-1}), \dots, (A_m + A_{-m}))$. These are multiplied respectively by $(K_0, K_{\frac{1}{2}}, \dots, K_{m/2})$ and the results added together.

The final value of $K_{n_0}(x)$ is obtained by successive approximation, 3, 5, 7... point formulæ being applied until consecutive results agree to the required number of figures.

In the following table, the Lagrange coefficients, as used for the computation of $K_{\frac{1}{2}}(x)$, $K_1(x)$ are given :—

Lagrange coefficients for computation of $K_{\frac{1}{2}}, K_1$ from $K_0, K_{\frac{1}{2}}, K_1, \dots$.		
Formula.	$K_{\frac{1}{2}}$ coefficients.	K_1 coefficients.
3 point	(3, 1)/4	(-5, 9)/4
5 point	(45, 20, -1)/64	(-35, 84, 15)/64
7 point	(350, 175, -14, 1)/512	(-210, 567, 162, -7)/512
9 point	(11025, 5880, -588, 72, -5), 16384	(-5775, 16632, 5940, -440, 27), 16384
11 point	(87318, 48510, -5544, 891, --110, 7)/131072	(-42042, 126126, 51480, -5005, 546, -33)/131072

The derivatives were obtained by using the relation

$$xK_v'(x) = -vK_v(x) - xK_{v-1}(x).$$

whence

$$K_{\frac{1}{2}}'(x) = -\frac{1}{4x}K_{\frac{1}{2}}(x) - K_{\frac{3}{2}}(x),$$

$$K_{\frac{3}{2}}'(x) = -\frac{3}{4x}K_{\frac{3}{2}}(x) - K_{\frac{5}{2}}(x).$$

Formulae for the higher derivatives follow by differentiation.

Tables of Bessel Functions : $K_v(x)$.

TABLE I.

Functions of order $1/2$, $3/2$, $5/2$.

x .	$K_{1/2}(x)$.	$K_{3/2}(x)$.	$K_{5/2}(x)$.	x .
0.1	3.5862	39.448	1187.0	0.1
0.2	2.2945	13.767	208.80	0.2
0.3	1.6952	7.3457	75.152	0.3
0.4	1.3283	4.6492	36.198	0.4
0.5	1.0750	3.2251	20.426	0.5
0.6	0.88799	2.3680	12.728	0.6
0.7	0.74388	1.8066	8.4863	0.7
0.8	0.62962	1.4166	5.9421	0.8
0.9	0.53712	1.1339	4.3169	0.9
1.0	0.46107	0.92214	3.2275	1.0
1.1	0.39778	0.75939	2.4688	1.1
1.2	0.34460	0.63177	1.9240	1.2
1.3	0.29957	0.53002	1.5227	1.3
1.4	0.26121	0.44778	1.2207	1.4
1.5	0.22834	0.38056	0.98945	1.5
1.6	0.20005	0.32507	0.80956	1.6
1.7	0.17560	0.27890	0.66778	1.7
1.8	0.15442	0.24020	0.55476	1.8
1.9	0.13600	0.20757	0.46374	1.9
2.0	0.11994	0.17991	0.38980	2.0
2.1	0.10591	0.15634	0.32925	2.1
2.2	0.093627	0.13618	0.27933	2.2
2.3	0.082855	0.11888	0.23791	2.3
2.4	0.073392	0.10397	0.20336	2.4
2.5	0.065066	0.091092	0.17438	2.5
2.6	0.057731	0.079935	0.14996	2.6
2.7	0.051261	0.070246	0.12931	2.7
2.8	0.045547	0.061813	0.11178	2.8
2.9	0.040496	0.054459	0.096833	2.9
3.0	0.036026	0.048035	0.084061	3.0
3.1	0.032068	0.042412	0.073111	3.1
3.2	0.028559	0.037484	0.063700	3.2
3.3	0.025447	0.033158	0.055590	3.3
3.4	0.022684	0.029356	0.048586	3.4
3.5	0.020230	0.026010	0.042524	3.5

TABLE I. (cont.).

x .	$K_{1/2}(x)$.	$K_{3/2}(x)$.	$K_{5/2}(x)$.	x .
3.6	0.018049	0.023062	0.037267	3.6
3.7	0.016109	0.020463	0.032701	3.7
3.8	0.014383	0.018168	0.028726	3.8
3.9	0.012846	0.016140	0.025262	3.9
4.0	0.011478	0.014347	0.022238	4.0
4.1	0.010258	0.012760	0.019594	4.1
4.2	0.0091706	0.011354	0.017281	4.2
4.3	0.0082009	0.010108	0.015253	4.3
4.4	0.0073356	0.0090028	0.013474	4.4
4.5	0.0065634	0.0080219	0.011911	4.5
4.6	0.0058739	0.0071508	0.010537	4.6
4.7	0.0052581	0.0063768	0.0093284	4.7
4.8	0.0047079	0.0056887	0.0082633	4.8
4.9	0.0042162	0.0050766	0.0073243	4.9
5.0	0.0037766	0.0045319	0.0064958	5.0
6.0	0.0012683	0.0014797	0.0020081	6.0
8.0	0.00014865	0.00016723	0.00021136	8.0
10.0	0.000017993	0.000019793	0.000023931	10.0

Tables of Bessel Functions : $K_\nu(x)$ and $K'_\nu(x)$.

TABLE II.

Functions and derivatives of order $1/4$, $3/4$.

x .	$K_{1/4}(x)$.	$K_{3/4}(x)$.	$-K'_{1/4}(x)$.	$-K'_{3/4}(x)$.	x .
0.1	2.69	5.6	12.3	45.	0.1
0.2	1.878	3.15	5.50	13.7	0.2
0.3	1.448	2.18	3.39	6.90	0.3
0.4	1.1651	1.642	2.370	4.244	0.4
0.5	0.9603	1.292	1.772	2.898	0.5
0.6	0.80399	1.0445	1.3795	2.1096	0.6
0.7	0.68058	0.8606	1.1037	1.6026	0.7
0.8	0.58086	0.71874	0.90026	1.2547	0.8
0.9	0.49893	0.60643	0.74502	1.0043	0.9
1.0	0.43074	0.51577	0.62346	0.81757	1.0
1.1	0.37342	0.44148	0.52635	0.67443	1.1
1.2	0.32486	0.37986	0.44754	0.56228	1.2
1.3	0.28345	0.32827	0.38278	0.47283	1.3
1.4	0.24794	0.28473	0.32900	0.40047	1.4
1.5	0.21736	0.24774	0.28396	0.34123	1.5
1.6	0.19091	0.21614	0.24597	0.29223	1.6
1.7	0.16797	0.18902	0.21372	0.25137	1.7
1.8	0.14801	0.16565	0.18621	0.21703	1.8
1.9	0.13060	0.14544	0.16262	0.18801	1.9
2.0	0.11538	0.12790	0.14233	0.16334	2.0

TABLE II. (*cont.*).

x .	$K_{\frac{1}{2}}(x)$.	$K_{\frac{3}{2}}(x)$.	$-K_{\frac{1}{2}}'(x)$.	$-K_{\frac{3}{2}}'(x)$.	x .
2.1	0.10204	0.11265	0.12480	0.14228	2.1
2.2	0.090341	0.099348	0.10961	0.12421	2.2
2.3	0.080054	0.087724	0.096426	0.10866	2.3
2.4	0.070999	0.077545	0.084941	0.095232	2.4
2.5	0.063017	0.068618	0.074919	0.083603	2.5
2.6	0.055973	0.060774	0.066156	0.073504	2.6
2.7	0.049750	0.053873	0.058480	0.064715	2.7
2.8	0.044246	0.047794	0.051744	0.057048	2.8
2.9	0.039374	0.042431	0.045826	0.050347	2.9
3.0	0.035057	0.037696	0.040618	0.044481	3.0
3.1	0.031229	0.033511	0.036030	0.039337	3.1
3.2	0.027833	0.029808	0.031983	0.034819	3.2
3.3	0.024817	0.026529	0.028409	0.030847	3.3
3.4	0.022137	0.023623	0.025251	0.027349	3.4
3.5	0.019755	0.021046	0.022457	0.024265	3.5
3.6	0.017635	0.018758	0.019983	0.021543	3.6
3.7	0.015749	0.016726	0.017791	0.019140	3.7
3.8	0.014069	0.014921	0.015847	0.017014	3.8
3.9	0.012572	0.013315	0.014121	0.015133	3.9
4.0	0.011238	0.011887	0.012589	0.013467	4.0
4.1	0.010049	0.010616	0.011228	0.011991	4.1
4.2	0.0089877	0.0094834	0.010018	0.010681	4.2
4.3	0.0080407	0.0084745	0.0089420	0.0095188	4.3
4.4	0.0071953	0.0075753	0.0079841	0.0085434	4.4
4.5	0.0064404	0.0067734	0.0071312	0.0075693	4.5
4.6	0.0057660	0.0060580	0.0063714	0.0067537	4.6
4.7	0.0051633	0.0054196	0.0056943	0.0060282	4.7
4.8	0.0046247	0.0048497	0.0050906	0.0053824	4.8
4.9	0.0041430	0.0043408	0.0045522	0.0048078	4.9
5.0	0.0037123	0.0038862	0.0040718	0.0042952	5.0
6.0	0.0012500	0.0012993	0.0013514	0.0014124	6.0
8.0	0.00014701	0.00015141	0.00015601	0.00016121	8.0
10.0	0.000017833	0.000018264	0.000018710	0.000019203	10.0

The authors' thanks are due to Dr. J. C. P. Miller, of Liverpool University, for his advice on the choice of calculation methods, and for the provision of facilities for carrying out the numerical computations. They also wish to thank British Insulated Cables, Ltd., Prescott, for permission to publish these results.

Bibliography.

- (1) See, for instance, H. S. Carslaw and J. C. Jaeger, 'Operational Methods in Applied Mathematics.'
- (2) G. N. Watson, 'Theory of Bessel Functions,' pp. 171-172.
- (3) G. N. Watson, 'Theory of Bessel Functions,' p. 737.

XCV. *The Field between Equal Semi-infinite Rectangular Electrodes or Magnetic Pole-pieces.*

By N. DAVY, University College, Nottingham *.

[Received July 26, 1944.]

Abstract.

The conformal transformation which gives the properties of the two-dimensional electrostatic system of two equal semi-infinite rectangular electrodes of opposite potentials, with a gap between them, is obtained, in terms of elliptic functions. Some equipotentials are mapped. Surface densities at various points and total charges on certain lengths of the conductors are calculated. Capacities of various parts of the system regarded as condensers are obtained, along with other electrical quantities. An experimental determination of equipotential curves made to test the theory is described.

§ 1. *Introduction.*

THE two-dimensional problem of finding the properties of the electrostatic field between two equal semi-infinite rectangular electrodes, with a gap between them, is commenced in Riemann-Weber's 'Differentialgleichungen der Physik' (Brunswick, 1927), pt. ii. pp. 303, *et seq.* The discussion stops when the differential equation of the transformation and the modulus of the elliptic functions to be used are obtained. Any reader wishing to obtain numerical results will find that much remains to be done thereafter. In the present paper the differential equation is integrated and a number of consequences are deduced from the solution. Experimental tests of the theory are also described.

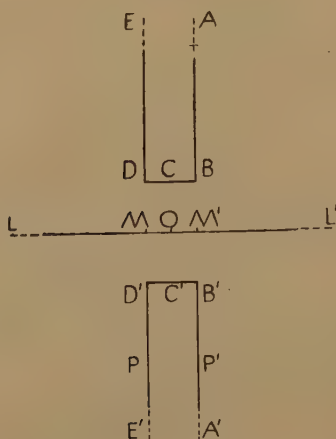
§ 2. *Transformation from z -plane to t -plane.*

The system is symmetrical and consists of two semi-infinite rectangular electrodes whose normal sections are represented by ABCDE and A'B'C'D'E' in fig. 1. Thus there is symmetry about the straight line LOL' and also about a normal to it through O (not drawn in). Let the potential of electrode ABCDE be $+\pi$ e.s.u. and of A'B'C'D'E' $-\pi$ e.s.u. of potential. Let the width of each electrode DB or D'B' be $2b$ cm. and of the gap DD' or BB' be $2a$ cm. The figure lies in the so-called z -plane, with the origin O at the central point of the gap. Hence the co-ordinates of B are $+b, +ia$, of D, $-b, +ia$, of B', $+b, -ia$, and of D', $-b, -ia$.

* Communicated by Professor L. F. Bates, D.Sc.

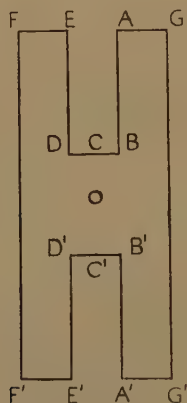
The transformation necessary to solve the problem is obtained by using the theorem of Schwarz and Christoffel. First, the theorem is applied to the letter H-shaped polygon, shown in fig. 2. The immediate problem

Fig. 1.



(z -plane). $BD=2$ cm., $DD'=4$ cm.

Fig. 2.



(z -plane).

Fig. 3.



(t -plane).

is to find a transformation changing the edge $GABDEFF'E'D'B'A'G'G$ into the real axis of another plane called the t -plane (fig. 3). Let the following values of t be allocated to points in the t -plane corresponding

to the corners of the polygon. Corresponding to G let $t=t_1$, to A, $t=t_2$, to B, $t=t_3$, to D, $t=t_4$, to E, $t=t_5$, and to F, $t=t_6$. Corresponding to the corners with dashed letters G', A', B', C', D', E' and F', the values of t may be taken as equal to those at G, A, B, C, D, E, and F, respectively, but opposite in sign, on account of the symmetry of the figure. Applying the theorem in the usual manner we get

$$dz = \frac{S(t^2-t_3^2)^{\frac{1}{2}}(t^2-t_4^2)^{\frac{1}{2}}dt}{(t^2-t_1^2)^{\frac{1}{2}}(t^2-t_2^2)^{\frac{1}{2}}(t^2-t_5^2)^{\frac{1}{2}}(t^2-t_6^2)^{\frac{1}{2}}} \dots \dots \dots (1)$$

where S is a constant. Corresponding to A let $t_2=0$, corresponding to B, let $t_3=-k$, to D let $t_4=-1$, and to E let $t_5=-\infty$. Now let the points G, A, E, F, G', A', E' and F' recede to an infinite distance from O. A and G coalesce, and so do E and F, E' and F', A' and G'. All this makes

$$dz = \frac{S(t^2-k^2)^{\frac{1}{2}}(t^2-1)^{\frac{1}{2}}dt}{t^2} \dots \dots \dots (2)$$

To integrate this, write $t = \frac{1}{\operatorname{sn} u}$, making k the modulus of the elliptic function used. As yet, the value of k is undetermined except that it lies between 0 and +1.

On substituting it is found that

$$\begin{aligned} dz &= -\frac{S \operatorname{dn}^2 u \operatorname{cn}^2 u du}{\operatorname{sn}^2 u} \dots \dots \dots (3) \\ &= -S \left\{ \frac{1}{\operatorname{sn}^2 u} - (1+k^2) + k^2 \operatorname{sn}^2 u \right\} du. \end{aligned}$$

The integral is

$$z = -S \left\{ u \left(k'^2 - \frac{2E}{K} \right) - 2Z(u) - \frac{\operatorname{cn} u \operatorname{dn} u}{\operatorname{sn} u} \right\} + T,$$

where $k'^2=1-k^2$. K and E are the complete elliptic integrals of the first and second kind, respectively. $Z(u)$ is the Jacobian elliptic function usually designated thus, and T is a constant of integration. By substituting the values of the known co-ordinates of the points B, D, B' and D', it is easily found that

$$\left. \begin{aligned} b &= \frac{a(K'k'^2 - 2K' + 2E')}{2(Kk'^2 - 2E)} \\ -S &= \frac{ai}{2E - k'^2K}, \quad T = -b. \end{aligned} \right\} \dots \dots \dots (4)$$

and

Hence the full transformation from z -plane to t -plane is

$$z = x + iy = \frac{ai \left\{ u \left(k'^2 - \frac{2E}{K} \right) - 2Z(u) - \frac{\operatorname{cn} u \operatorname{dn} u}{\operatorname{sn} u} \right\}}{2E - k'^2K} - b. \dots \dots (5)$$

K' is the same function of k' as K is of k , and E' is the same function of k' as E is of k . A table of values of t , $1/t = \operatorname{sn} u$ and u at various points is inserted here.

TABLE I.

Point.	$t.$	$\frac{1}{t} = \text{sn } u.$	$u.$
A.....	0	∞	$-iK'$
B.....	$-k$	$-\frac{1}{k}$	$-(K+iK')$
C	$-\sqrt{k}$	$-\frac{1}{\sqrt{k}}$	$-\left(K+\frac{iK'}{2}\right)$
D	-1	-1	$-K$
E.....	$-\infty$	0	0

At A', B', C', D' and E' the values are equal in magnitude but opposite in sign to those at A, B, C, D and E respectively. Also in equation (5) u is generally complex and can be written

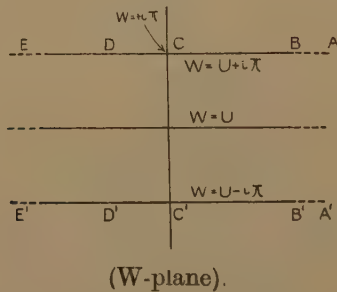
$$u=p+iq,$$

where p and q are real.

§ 3. Transformation from W -plane to t -plane.

So far, the discussion has been purely on geometrical lines. It is now necessary to link it up with the electrical conditions. The upper and lower electrodes in fig. 1 have potentials $+\pi$ and $-\pi$ e.s.u. respectively. The potential of the central plane, if a conductor, represented by the x -axis, *i. e.*, the line LOL', is zero. As usual in such problems we consider a so-called W -plane (fig. 4) in which there are two horizontal

Fig. 4.



lines, the first, ABCDE, at a height π cm. above the real axis, the other, A'B'C'D'E', at a depth π cm. below the real axis. Here $W=U+iV$, where U is the abscissa and iV the ordinate of any point. At any point on the straight line AE, $W=U+i\pi$. At any point on A'E', $W=U-i\pi$, and along the real axis $W=U$, at any point. U is a variable in each case. Hence AE, in fig. 4, corresponds to the upper electrode in fig. 1, and A'E' to the lower one. In fig. 4, AEE'A'A is regarded as a polygon

whose boundary is transformed to the real axis in the t -plane, by the Schwarz and Christoffel theorem, which gives

$$dW = \alpha \frac{dt}{t}, \quad \dots \dots \dots (6)$$

where α is a constant. Hence

$$W = \alpha \log t + \beta. \quad \dots \dots \dots (7)$$

Since W changes by $-2\pi i$ when t changes by $+e^{i\pi}$, in going from A' to A , $\alpha = -2$.

Also, if we make the point C in the W -plane, where $W = i\pi$, correspond to a point C in the t -plane, where $t = -\sqrt{k}$, and also to the point C in the z -plane, we get

$$\beta = i\pi + 2 \log (-\sqrt{k}).$$

Hence
$$W = -2 \log t + i\pi + 2 \log (-\sqrt{k}).$$

Now
$$e^{i\pi} = \cos \pi + i \sin \pi = -1,$$

and
$$e^{-i\pi} = \cos \pi - i \sin \pi = -1.$$

Hence
$$\log (-1) = \pm i\pi$$

and
$$\log i^2 = \pm i\pi \quad \text{and} \quad 2 \log i = \pm i\pi.$$

Now
$$t = \frac{1}{\operatorname{sn} u}$$

hence
$$W = 2 \log \operatorname{sn} u + 2 \log i + 2 \log (-\sqrt{k}) \quad \dots \dots \dots (8)$$

$$= 2 \log \operatorname{sn} u + 2 \log (-i\sqrt{k})$$

$$= 2 \log \operatorname{sn} u + \log (-k). \quad \dots \dots \dots (9)$$

$$W = U + iV = \log (-k \operatorname{sn}^2 u). \quad \dots \dots \dots (10)$$

Write
$$u = p + iq$$

$$\operatorname{sn} u = \operatorname{sn} (p + iq)$$

$$= \frac{s_1 d_2 + i c_1 d_1 s_2 c_2}{c_2^2 + k^2 s_1^2 s_2^2}, \quad \dots \dots \dots (11)$$

where $s_1 = \operatorname{sn} p$, $c_1 = \operatorname{cn} p$, $d_1 = \operatorname{dn} p$, all to modulus k , and $s_2 = \operatorname{sn} q$, $c_2 = \operatorname{cn} q$, $d_2 = \operatorname{dn} q$, all to modulus k' .

We may also write

$$\operatorname{sn} u = \operatorname{Re} e^{i\phi},$$

where

$$\left. \begin{aligned} R \sin \phi &= \frac{c_1 d_1 s_2 c_2}{c_2^2 + k^2 s_1^2 s_2^2} \\ \text{and} \quad R \cos \phi &= \frac{s_1 d_2}{c_2^2 + k^2 s_1^2 s_2^2} \end{aligned} \right\} \dots \dots \dots (12)$$

Also R is real, and
$$\tan \phi = \frac{c_1 d_1 s_2 c_2}{s_1 d_2}. \quad \dots \dots \dots (13)$$

Hence, by (9), $W=2 \log \operatorname{sn} u + \log (-k)$,

$$=2 \log \operatorname{Re}^{\phi} + \log (+k) + \log (-1),$$

and $W=U+iV$

$$=2 \log R+2i\phi+\log k \pm i\pi.$$

Hence

$$\left. \begin{aligned} U &= 2 \log R + \log k = \log k R^2, \\ V &= 2\phi \pm \pi. \end{aligned} \right\} \dots \dots \dots (14)$$

We take the solution $V=2\phi-\pi$ here, so that when $\phi=\pi$, $V=\pi$, corresponding to the upper electrode in the z -plane. When $\phi=\frac{\pi}{2}$, $V=0$, corresponding to the central plane, and when $\phi=0$, $V=-\pi$, corresponding to the lower electrode. By (14) equipotential curves are obtained by putting ϕ equal to a constant, or more conveniently

$$\tan \phi = \frac{c_1 d_1 s_2 c_2}{s_1 d_2} = H_1, \text{ a constant. } \dots \dots \dots (15)$$

Lines of force are given by

$$R = \text{a constant} \dots \dots \dots (16)$$

Equipotentials can be drawn only after the moduli k and $k'=(1-k^2)^{\frac{1}{2}}$ have been found numerically.

§ 4. *Determination of the Numerical Values of the Moduli k and k' .*

By equation (4)

$$\frac{b}{a} = \frac{K'k'^2 - 2K' + 2E'}{2(Kk'^2 - 2E)}.$$

Let $a=nb$, then $n(K'k'^2 - 2K' + 2E') = 2(Kk'^2 - 2E)$

and

$$k'^2 = \frac{2n(K' - E') - 4E}{nK' - 2K} \dots \dots \dots (17)$$

To use this a definite ratio of gap width to electrode width in fig. 1 must be selected. The special case selected here is $n=2$, $a=2b$. On substituting this in (17) we get

$$k'^2 = \frac{2\{K' - E' - E\}}{K' - K} \dots \dots \dots (18)$$

Since $k'^2=1-k^2$, $k'=\cos \theta$ if $k=\sin \theta$, where θ is the modular angle.

To find the numerical value of k' two graphs are plotted on the same sheet. For the first plot $x=\theta$, $y=\cos^2 \theta$, obtained from ordinary tables, and to get the second graph plot

$$x=\theta, \quad y = \frac{2(K' - E' - E)}{K' - K},$$

using values of E , E' , K , and K' , and the corresponding values of the modular angle θ obtained from pp. 261 *et seq.* of the Smithsonian 'Mathematical Formulæ and Tables of Elliptic Functions' (Publication 2672 of the Smithsonian Institution), dated 1922, by E. P. Adams and R. L.

Hippisley. These graphs intersect when the value of the abscissa θ is very nearly 6.5° , so that $k'=\cos \theta=0.9936$ and $k=\sin \theta=0.1132$ approx. The ratio of k' to k is 8.777. There are no tables for $\theta=6.5^\circ$ itself.

§5. *To Draw the Equipotential Curves.*

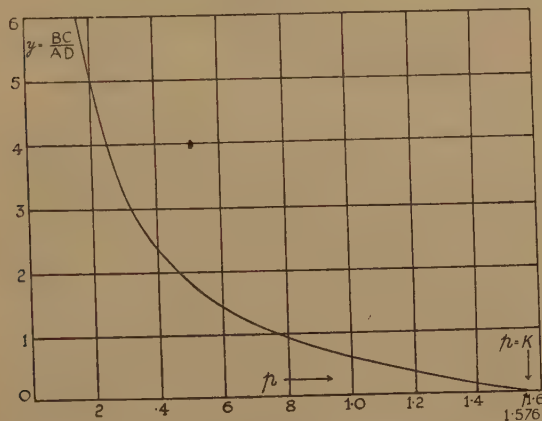
By (15)

$$\tan \phi = \frac{c_1 d_1 s_2 c_2}{s_1 d_2} = H_1.$$

Here then s_1 , c_1 and d_1 have a modulus $k=0.1132$ and s_2 , c_2 and d_2 have a modulus $k'=0.9936$. Using the system of symbols on p. 248 of the Smithsonian Tables, we write

$$\frac{c_1 d_1}{s_1} = \frac{BCK'}{DA} \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Fig. 5.



A, B, C and D ought to be determined using a modulus $k=\sin 6.5^\circ=0.1132$, but their values for that modulus are not tabulated. Their values for a modulus $k=\sin 5^\circ$, tabulated on pp. 260-261 of the Smithsonian Tables, only differ slightly from those for $k=\sin 6.5^\circ$. The full expression $\frac{BCK'}{DA}$ for modulus $k=\sin 5^\circ$ differs by much less than one per cent. from its value for modulus $k=\sin 6.5^\circ$. Hence we use the tables for $k=\sin 5^\circ$ to get numerical values of $\frac{BC}{DA}$ and we plot a graph (fig. 5) in which $x=F$, the quantity in the second column on p. 260, and $y=\frac{BC}{DA}$. Also the complex variable $u=p+iq$, and the F just mentioned is the p of $u=p+iq$. We also need values of $\frac{s_2 c_2}{d_2}$ in equation (15).

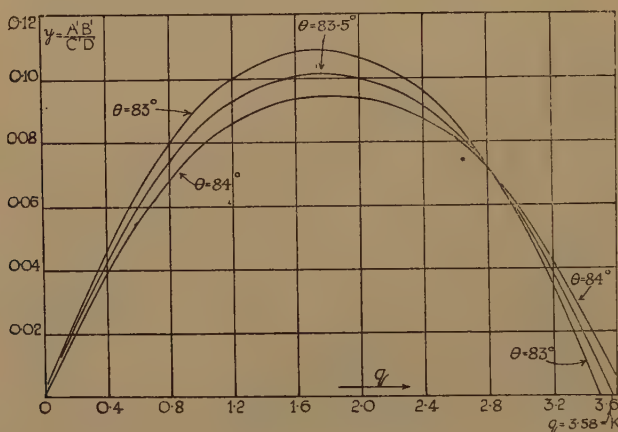
For exact calculations the modulus of s_2 , c_2 and d_2 should be $\sin 83.5^\circ=0.9936$, but tables for this modulus are not available. The Smithsonian Tables do, however, gives values of functions for moduli

equal to $\sin 83^\circ$ and $\sin 84^\circ$ respectively on pp. 296–299. Since the values change rapidly in the neighbourhood of these modular angles, we must interpolate graphically to get values corresponding to a modulus equal to $\sin 83.5^\circ$. When in the expression $\frac{s_2 c_2}{d_2}$ we replace s_2 , c_2 and d_2 by the symbols of page 248, we get

$$\frac{s_2 c_2}{d_2} = \frac{A'B'}{kC'D'} \quad \dots \dots \dots (20)$$

Values of $\frac{A'B'}{C'D'}$ are calculated for both $\theta=83^\circ$ and $\theta=84^\circ$ and two graphs are plotted on the same sheet (fig. 6), viz., $y_1 = \frac{A'B'}{C'D'}$, $x=F'$, using the

Fig. 6.



values of A' , B' , C' , D' and F' on pp. 296–297, and also $y_2 = \frac{A'B'}{C'D'}$, $x=F'$, using the values on pp. 298–299. The F' used here is also the q of $u=p+iq$. After drawing the two graphs (fig. 6), points with ordinates midway between them, i. e., $y = \frac{y_1 + y_2}{2}$, are marked on the paper. The graph passing through these points represents the graph $x=q$, $y = \frac{A'B'}{C'D'}$, when the modulus is equal to $\sin 83.5^\circ = 0.9936$, with fair accuracy (fig. 6). From fig. 5 and fig. 6 we pick out abscissæ p and q , such that the product of the corresponding ordinates $y = \frac{BC}{DA}$ and $y = \frac{A'B'}{C'D'}$ is equal to a constant H_2 , say.

$$\frac{BCA'B'}{DAC'D'} = H_2 \quad \dots \dots \dots (21)$$

Now, by (13),

$$\tan \phi = \frac{c_1 d_1 s_2 c_2}{s_1 d_2} = \frac{BCK'A'B'}{DAC'D'k} = H_2 \frac{k'}{k} = H_1 \text{ say.} \quad \dots \dots (22)$$

The values of p and q thus obtained correspond, therefore, to points on equipotential curves and are to be used in substituting $u=p+iq$ in equation (5), in order to find the co-ordinates of points on equipotentials. To do this we first insert in equation (5) the values

$$\operatorname{sn} u = u - \frac{(k^2+1)}{3}u^3 + \frac{(k^4+2k^2+4)}{5}u^5 \text{ approx.} \quad (23)$$

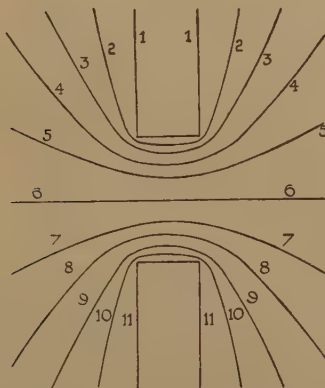
$$\operatorname{cn} u = 1 - \frac{u^2}{2} + \frac{(1+4k^2)u^4}{4} \text{ approx.} \quad (24)$$

$$\operatorname{dn} u = 1 - \frac{k^2u^2}{2} + \frac{k^2(4+k^2)u^4}{4} \text{ approx.} \quad (25)$$

$$Z(u) = \left(1 - \frac{E}{U}\right)u - \frac{2k^2u^3}{3} + \frac{8k^2(k^2+1)u^5}{5} \text{ approx.} \quad (26)$$

Since $k = \sin 6.5^\circ = 0.113$, $k^2 = 0.0128$, terms in k^4 may be neglected, provided u is less than unity, if we only seek accuracy to about one part

Fig. 7.



in a thousand. Next write $u=p+iq$ and $z=x+iy$ in (5) and find x and y separately. After some labour, we get

$$x = -b + \frac{a}{2E - Kk'^2} \left\{ -\frac{q}{p^2 + q^2} + \frac{2q}{3} - \frac{q(3p^2 - q^2)}{45} - \frac{2q(5p^4 - 10p^2q^2 + q^4)}{945} \right\},$$

$$y = \frac{a}{2E - Kk'^2} \left\{ -\frac{p}{p^2 + q^2} - \frac{2p}{3} + \frac{p(p^2 - 3q^2)}{45} + \frac{2p(p^4 - 10p^2q^2 + 5q^4)}{945} \right\}. \quad (27)$$

These values of x and y represent points on equipotentials, not only in the present case in which $a=2b$, but in any case of equal symmetrical semi-infinite rectangular electrodes in which k is so small compared with unity that equations (23), (24), (25) and (26) are valid. When $a=2$ cm., $b=1$ cm., $k=0.113$, and $\frac{a}{2E - Kk'^2} = +1.270$ app. Tables of values of p , q , x and y of equipotential curves, worked out in the above manner for the cases in which the constant H_2 of equation (21) is equal

to ± 0.025 , ± 0.05 , ± 0.1 and ± 0.2 respectively, are now given. The corresponding graphs are shown in fig. 7. If the potential of the upper electrode, the central plane and the lower plane were $+100$ volts, zero and -100 volts respectively, the potentials of these curves would be ± 86.3 volts, ± 73.6 volts, ± 54.1 volts and ± 33.0 volts, respectively.

TABLE II.

$$x \text{ and } y \text{ in cm. } H_2 = 0.025. \quad H_1 = H_2 \frac{k'}{k} = \tan \phi = 0.219.$$

$$V = -0.863\pi \text{ e.s.u.} = -2.71 \text{ e.s.u.}$$

p	1.32	1.10	0.78	0.47	0.32	0.25
q	1.54	0.50	0.23	0.11	0.080	0.060
x	-0.29	-1.06	-1.27	-1.55	-1.88	-2.10
y	-1.73	-1.87	-2.15	-2.99	-4.03	-5.02

TABLE III.

$$x \text{ and } y \text{ in cm. } H_2 = 0.05. \quad H_1 = H_2 \frac{k'}{k} = \tan \phi = 0.439.$$

$$V = -0.736\pi = -2.314 \text{ e.s.u.}$$

p	1.10	0.78	0.58	0.47	0.33
q	1.54	0.50	0.30	0.22	0.15
x	-0.22	-1.34	-1.65	-1.86	-2.32
y	-1.50	-1.82	-2.20	-2.61	-3.48

TABLE IV.

$$x \text{ and } y \text{ in cm. } H_2 = 0.10. \quad H_1 = H_2 \frac{k'}{k} = \tan \phi = 0.878.$$

$$V = -0.541\pi = -1.701 \text{ e.s.u.}$$

p	0.78	0.74	0.69	0.65	0.47	0.38	0.33	0.25	0.17
q	1.54	1.14	0.95	0.84	0.50	0.38	0.30	0.22	0.14
x	-0.29	-0.85	-1.09	-1.24	-1.93	-2.36	-2.65	-3.37	-4.52
y	-1.13	-1.20	-1.27	-1.30	-1.68	-2.00	-2.38	-3.07	-4.55

On account of symmetry the labour of calculating co-ordinates is only required for one-quarter of each curve. The edges of the upper and lower electrodes both correspond to $H_2 = 0$, zero in the case of the upper one being approached via negative values of H_2 , and via positive values of H_2 in the case of the lower electrode. This statement is to be checked by equation (22) rather than (27). See also Table X. at the end of the paper, for details of fig. 7.

These equipotentials, so awkward to obtain, form a family of curves which may, perhaps, be described as transcendental or "bull-nosed" hyperbolas, since they degenerate into a set of confocal hyperbolas when the electrodes become infinitely narrow. In hydrodynamics, such curves would be stream lines through a gap of the given shape.

TABLE V.

$$x \text{ and } y \text{ in cm. } H_2=0.20. \quad H_1=H_2 \frac{k'}{k}=\tan \phi=1.755.$$

$$V=-0.330\pi=-1.04 \text{ e.s.u.}$$

p	0.205	0.255	0.325	0.470	0.170	0.140	0.120
q	0.39	0.50	0.70	1.54	0.305	0.26	0.22
x	-3.22	-2.80	-1.89	-0.529	-3.92	-4.57	-5.26
y	-1.52	-1.36	-0.98	-0.716	-1.91	-2.16	-2.53

§ 6. *Lines of Force.*

When the equipotentials have been drawn as in fig. 7 the lines of force, which are orthogonal to them, may be obtained by the optical method of placing a plane mirror with a straight edge, so as to stand vertically on the paper, the straight edge forming the horizontal edge on which the mirror rests. The same straight lower edge is set at right angles to any equipotential at any point, first as gauged by the eye directly and then, further, as a finer adjustment, by looking into the mirror and rotating slightly as necessary until the image of the equipotential concerned appears to be continuous with the real curve in front of the mirror. Then a short pencil-line ruled along a bit of the edge of the mirror will be an element of a line of force. Other elements obtained thus can be drawn and joined smoothly to form a continuous curve. The lines of force form a family of curves which degenerate into confocal ellipses when the width of the electrodes becomes zero. An algebraic method of calculating points on lines of force analogous to the one for equipotentials has been worked out, using equations (16) and (12)*.

§ 7. *Surface Densities on the Electrodes.*

In all transformations from the z -plane to the W -plane the electric intensity or field strength at any point is

$$F=\left|\frac{dW}{dz}\right| \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

where the vertical lines mean that F is the modulus of $\frac{dW}{dz}$. (See Jeans' 'Electricity and Magnetism,' 4th ed. (1920), p. 265.)

Hence, on the surface of conductors, the surface density of charge

$$\sigma=\frac{1}{4\pi}\left|\frac{dW}{dz}\right| \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

* This was done by Mr. A. R. Maxwell.

Since, by equation (3),

dz = - \frac{S \operatorname{dn}^2 u \operatorname{cn}^2 u \operatorname{dn} u}{\operatorname{sn}^2 u},

and by (10),

W = \log (-k \operatorname{sn}^2 u)

and

dW = \frac{2 \operatorname{cn} u \operatorname{dn} u \operatorname{dn} u}{\operatorname{sn} u},

we soon get

\sigma = \frac{1}{4\pi} \left| \frac{dW}{dz} \right| = \left| \frac{-i(2E - k'^2 K) \operatorname{sn} u}{2\pi a \operatorname{cn} u \operatorname{dn} u} \right|. \tag{30}

TABLE VI.

Position of the point.	p .	iq .
A	0	$-iK'$
From A to B.....	$0 > p > -K$	$-iK'$
B	$-K$	$-iK'$
From B to C	$-K$	$-iK' < iq < -\frac{iK'}{2}$
C	$-K$	$-\frac{iK'}{2}$
From C to D.....	$-K$	$-\frac{iK'}{2} < iq < 0$
D	$-K$	0
From D to E.....	$-K < p < 0$	0
E	0	0
L	0	0
From L to O.....	0	$0 < iq < \frac{iK'}{2}$
O	0	$\frac{iK'}{2}$

Equation (30) is true for all values of k , *i. e.*, whatever the ratio of gap width to electrode width. As before, $u = p + iq$, and is sometimes real, sometimes complex. In order to facilitate the calculation of σ a table of values of p and iq at various points on the electrodes (see fig. 1) is inserted.

On account of the symmetry of the figure it will suffice, in calculating σ on the electrodes, to do so from E to D along one side, and from D to C on one end.

Along ED.—Here $-K < p < 0$, $q = 0$, $x = -b$ and, by (27),

y = +1.270 \left\{ -\frac{1}{p} - \frac{2p}{3} + \frac{p^3}{45} + \frac{2p^5}{945} \right\}.

To obtain numerical values of σ we make use of the functions A, B, C, D, defined on p. 248 of the Smithsonian Tables.

$$\begin{aligned}\sigma &= \left| \frac{-i(2E - k'^2K) \operatorname{sn} u}{2\pi a \operatorname{cn} u \operatorname{dn} u} \right| \\ &= \frac{(2E - k'^2K)A \cdot D}{2\pi a k' B \cdot C} \text{ in all cases,} \\ &= 1.44 \times 10^{-3} \frac{AD}{BC} \text{ e.s.u. per sq. cm. in the present case.}\end{aligned}$$

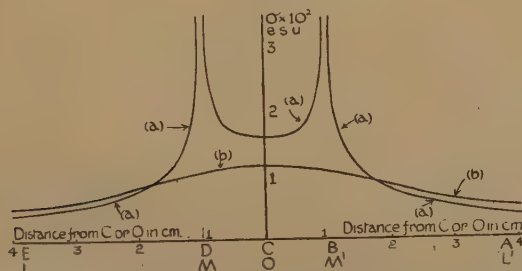
Moreover, $\frac{AD}{BC}$ is the reciprocal of any ordinate of the curve in fig. 5 whose abscissa p gives y as shown above.

TABLE VII.

Along ED, $q=0$, $a=2$, $b=1$, $k=0.113$. $x=-b=-1$ cm.

y cm.	∞	12.8	6.52	4.48	3.51	2.96	2.62	2.40	2.25	2.15	2.08
$\sigma \times 10^3$ e.s.u.	0	1.44	2.79	4.29	5.98	7.76	9.90	12.2	15.0	17.9	23.2

Fig. 8.



Surface density at points on (a) an electrode whose potential is π e.s.u., and (b) a central plane at zero potential.

At the corner D $x=-1.00$ cm., $y=2.00$ cm., $\sigma=\infty$. There is, therefore, a very sharp rise in σ close to the corner. The graph connecting σ and distance from C, *i. e.* $(y-1)$, is plotted in fig. 8. For convenience, the ordinates represent values of σ , and the abscissæ, $= (y-1) =$ distances from C measured along the edge around the corner. The parts of the graph concerned are the left and right wings of graph (a), fig. 8, for the edges DE and BA respectively.

Along CD.—Here

$$\begin{aligned}p &= -K, \quad -\frac{iK}{2} < iq < 0, \quad u = p + iq = -K + iq, \\ \operatorname{sn}(iq - K) &= \frac{\operatorname{cn} iq}{\operatorname{dn} iq}, \quad \operatorname{cn}(iq - K) = -k' \frac{\operatorname{sn} iq}{\operatorname{dn} iq}, \quad \operatorname{dn}(iq - K) = \frac{k'}{\operatorname{dn} iq}.\end{aligned}$$

$$\begin{aligned}\sigma &= \left| \frac{-i(2E - k'^2 K) \operatorname{sn} u}{2\pi a \operatorname{cn} u \operatorname{dn} u} \right| \\ &= \left| \frac{-i(2E - k'^2 K) \operatorname{cn} iq \operatorname{dn} iq}{2\pi a k'^2 \operatorname{sn} iq} \right| \\ &= \frac{(2E - k'^2 K) \operatorname{dn}(q, k')}{2\pi a \operatorname{cn}(q, k') \operatorname{sn}(q, k')} \dots \dots \dots (32)\end{aligned}$$

$$= \frac{(2E - k'^2 K) k C'D'}{2\pi a A'B'} \dots \dots \dots (33)$$

using the symbols of p. 248 of the Smithsonian Tables, but with the modulus k' . Inserting numerical values we get

$$\sigma = 1.63 \times 10^{-3} \frac{C'D'}{A'B'} \text{ e.s.u. per sq. cm. } \dots \dots \dots (34)$$

Also $\frac{C'D'}{A'B'}$ is the reciprocal of the ordinate of the middle graph in fig. 6.

TABLE VIII.

Along CD, $y=2 \text{ cm.}$ $p=-K$ and $-\frac{iK'}{2} < iq < 0$.

$q \dots \dots \dots$	0	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.79
$x \text{ cm. } \dots \dots$	-1.00	-0.916	-0.835	-0.718	-0.564	-0.379	-0.160	0
$\sigma \times 10^2 \text{ e.s.u.}$	∞	2.71	2.18	1.90	1.75	1.66	1.61	1.594

$1.79 = \frac{K'}{2}$ corresponding to C the mid-point of the end. At C itself

$$u = -\left(K + \frac{iK'}{2}\right), \quad \operatorname{sn} u = -\frac{1}{\sqrt{k}}, \quad \operatorname{cn} u = -i\sqrt{\frac{1-k}{k}}, \quad \operatorname{dn} u = \sqrt{1-k},$$

$$\text{and } \sigma = \left| \frac{-i(2E - k'^2 K) \operatorname{sn} u}{2\pi a \operatorname{cn} u \operatorname{dn} u} \right| = \frac{(2E - k'^2 K)}{2\pi a(1-k)} = 1.594 \times 10^{-2} \text{ e.s.u. } \quad (35)$$

The part of the graph concerned is the central part of graph (a), fig. 8, representing the end DCB.

Along LOL', the Central Plane.

Suppose the central plane in fig. 1, midway between the ends of the electrodes, to be a conductor of electricity. It is represented in section by the straight line LOL', *i. e.*, $y=0$. From L to O, $-\infty < x < 0$, $p=0$, iq is variable and $0 < iq < \frac{iK'}{2}$.

The surface density

$$\sigma = \left| \frac{-i(2E - k'^2 K) \operatorname{sn} u}{2\pi a \operatorname{cn} u \operatorname{dn} u} \right| \text{ by (30),}$$

$$\begin{aligned}
 &= \left| \frac{-i(2E - k'^2 K) \operatorname{sn} iq}{2\pi a \operatorname{cn} iq \operatorname{dn} iq} \right| \text{ using modulus } k, \\
 &= \left| \frac{(2E - k'^2 K) \operatorname{sn}(q, k') \operatorname{cn}(q, k')}{2\pi a \operatorname{dn}(q, k')} \right| \\
 &= \frac{(2E - k'^2 K) A' B'}{2\pi a k C' D'} \dots \dots \dots (36)
 \end{aligned}$$

where A', B', C', D' are the A, B, C, D of p. 248 of the Smithsonian Tables, but to modulus k' .

Hence
$$\sigma = \frac{0.111 A' B'}{C' D'} \text{ e.s.u. per sq. cm. } \dots \dots \dots (37)$$

The middle graph of fig. 6 is used to get $\frac{A' B'}{C' D'}$ for any value of q between zero and $\frac{K'}{2}$, and thus to calculate σ . At O itself, $q = \frac{K'}{2}$, $p = 0$, $u = \frac{iK'}{2}$.

Using equation (30) directly, substitute $\operatorname{sn} u = \frac{i}{\sqrt{k}}$, $\operatorname{cn} u = \sqrt{\frac{1+k}{k}}$, $\operatorname{dn} u = \sqrt{1+k}$, thus obtaining

$$\sigma \text{ at O} = \frac{(2E - k'^2 K)}{2\pi a(1+k)} \dots \dots \dots (38)$$

holding for all such systems of conductors, whatever the value of k , and in our special case when $k = 0.113$,

$$\sigma \text{ at O} = 1.13 \times 10^{-2} \text{ e.s.u. } \dots \dots \dots (39)$$

TABLE IX.

σ = Surface densities along OL. $y = 0$.

$q \dots \dots \dots$	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	$1.79 = \frac{K'}{2}$
$x \dots \dots \dots$	-13.63	-7.19	-3.84	-2.60	-1.90	-1.39	-1.00	-0.658	-0.350	0
$\sigma \times 10^2 \text{ e.s.u.} \dots$	0.130	0.254	0.475	0.664	0.818	0.951	1.03	1.08	1.12	1.128

Values of σ at points along LOL' are represented by ordinates of the graph (b) of fig. 8. There is naturally a peak value at O, the centre of symmetry.

In the magnetic system of two semi-infinite rectangular pole-pieces which are magnetic equipotentials, *i. e.*, of infinite permeability, most of the above theory would hold good. σ would be replaced by the intensity of magnetization. Its numerical values in e.m.u. would be the same as those calculated above for σ in e.s.u. The central plane would be a plane wall of infinite magnetic permeability.

§ 8. Total Charges on the Electrodes, and Capacities.

Every point in the system has its own $z=x+iy$. The total charge per cm. depth perpendicular to the plane of the figure, between points defined by parameters z_1 and z_2 , is

$$\begin{aligned} Q &= \int_{z_1}^{z_2} \sigma \cdot dz \\ &= \frac{1}{4\pi} \left| \int_{z_1}^{z_2} \frac{dW}{dz} \cdot dz \right| \\ &= \frac{1}{4\pi} |W_2 - W_1|. \quad \dots \dots \dots (40) \end{aligned}$$

Here $W = \log(-k \operatorname{sn}^2 u)$, by (10).

Hence
$$Q = \left| \frac{1}{4\pi} \log \frac{\operatorname{sn}^2 u_2}{\operatorname{sn}^2 u_1} \right| = \left| \frac{1}{2\pi} \log \frac{\operatorname{sn} u_2}{\operatorname{sn} u_1} \right|. \quad \dots \dots \dots (41)$$

If u_2 is the parameter of the point E in fig. 1, $u_2=0$, $\operatorname{sn} u_2=0$ and $\log \operatorname{sn} u_2 = -\infty$. If u_2 is the parameter of the point A in fig. 1, $u_2=-iK'$, $\operatorname{sn} u_2 = \infty$, $\log \operatorname{sn} u_2 = \infty$. The charge Q_1 per cm. depth perpendicular to the plane of the paper, residing between any point P not at infinity, and either E or A is, therefore, infinite. The charge Q_1 per cm. depth between E and A is also infinite. Similarly the charge on either half, OL or OL', of the central plane conductor, or on the whole LOL' is infinite. Since between the pair of rectangular electrodes there is a potential difference 2π e.s.u., the capacity of the condenser formed by them is $\frac{Q_1}{2\pi} = \infty$. So is the capacity between either electrode and the central plane, whose potential is zero.

For finite lengths of the edges of the electrodes, $Q_1 = \frac{1}{2\pi} \log \frac{\operatorname{sn} u_2}{\operatorname{sn} u_1}$ per cm. length, by (41), and is finite. For example, it is finite for a finite length of edge extending equally up each side of either electrode. Thus at the point C' in the middle of the end of the lower electrode $u = +K + \frac{iK'}{2}$ and $\operatorname{sn} u = +\frac{1}{\sqrt{k}}$. The charge Q_1 per cm. depth between C' and any point P on the edge of the lower electrode is

$$\begin{aligned} Q_1 &= \frac{1}{2\pi} \log \frac{\operatorname{sn} u_2}{\frac{1}{\sqrt{k}}} \\ &= \frac{1}{4\pi} \log (k \operatorname{sn}^2 u_2). \quad \dots \dots \dots (42) \end{aligned}$$

Between two points, P and P', at equal distances down the side of the lower electrode

$$\begin{aligned} Q_1 &= 2 \times \frac{1}{4\pi} \log (k \operatorname{sn}^2 u_2), \text{ by symmetry,} \\ &= \frac{1}{2\pi} \log (k \operatorname{sn}^2 u_2). \quad \dots \dots \dots (43) \end{aligned}$$

The capacity per cm. depth between the length PC'P' of the lower electrode and the corresponding portion of the upper one is

$$C_1 = \frac{Q_1}{2\pi} = \frac{1}{4\pi^2} \log(k \operatorname{sn}^2 u_2) \text{ e.s.u.} \quad (44)$$

From C', the mid-point of the end and D', one corner, the total charge is

$$Q = \frac{1}{2\pi} \log \frac{\operatorname{sn} u_D}{\operatorname{sn} u_C}, \text{ by (41).}$$

At D', $u=K$, $\operatorname{sn} K=1$, and at C', $u=K+\frac{iK'}{2}$, $\operatorname{sn} u = \frac{1}{\sqrt{k}}$.

Hence $Q_1 = \frac{1}{2\pi} \log \sqrt{k} = \frac{1}{4\pi} \log k, \quad (47)$

Between B' and D', the two corners, the charge is

$$Q_1 = \frac{1}{2\pi} \log k, \quad (48)$$

twice that from D' to C', by symmetry. Since there is an equal opposite charge on the opposite end and the p.d. is 2π , the capacity per cm. depth between the two ends alone is

$$C_1 = \frac{Q_1}{V} = \frac{1}{4\pi^2} \log k \text{ e.s.u.} \quad (49)$$

The charge per cm. depth along one side, say from D to P, is the difference of the charges on C'P and on C'D', *i. e.*, it is

$$Q_{DP} = \frac{1}{4\pi} \log \operatorname{sn}^2 u, \quad (50)$$

u being the parameter at P. It would appear from the above that the surface density at the corners of the ends is infinite, and also by geometry that the area of the corners is zero, yet the total charge on each corner is zero.

§ 9. *Electric Resistance or Magnetic Reluctance of the Gap.*

If the gap between the two electrodes is filled with a medium of electric resistivity ρ e.m.u., the resistance per cm. depth between PCP' and the corresponding part of the opposite electrode is, by an old theorem,

$$R_1 = \frac{\rho}{4\pi C_1} \text{ e.m.u.}, \quad (51)$$

where C_1 is the capacity per cm. depth in e.s.u. The portion of gap concerned is bounded laterally by the lines of flow linking points on one electrode to the opposite one. The lines of flow in the current problem coincide with the lines of force in the corresponding electrostatic or

magnetic cases. Now the capacity per cm. depth between PCP' and its counterpart is

$$C_1 = \frac{1}{4\pi^2} \log(k \operatorname{sn}^2 u_2), \text{ by (44).}$$

Hence $R_1 = \frac{\pi\rho}{\log(k \operatorname{sn}^2 u_2)}$ e.m.u. per cm. depth, (52)

u_2 being the parameter at P or P'. When P coincides with a corner, say D', then $u_2 = K$, $\operatorname{sn} u_2 = +1$, $\operatorname{sn}^2 u_2 = +1$, and -

$$R_1 = \frac{\pi\rho}{\log k} \text{ e.m.u. per cm. depth.} \quad (53)$$

The reluctance in the corresponding magnetic problem is obtained by replacing ρ by $\frac{1}{\mu}$, where μ is the permeability of the material in the gap. Between the central plane and either electrode the resistance or reluctance is half that given by (53) or (52) in corresponding circumstances.

§ 10. Mechanical Forces on Conductors.

The normal force on an element of area of any electrostatic conductor, when the dielectric in front of it has a constant equal to unity, is $2\pi\sigma^2$ dynes per sq. cm. In the type of system under discussion this force, which may be named

$$X_1 = 2\pi \left(\frac{2E - k'^2 K}{2\pi a} \right)^2 \frac{\operatorname{sn}^2 u}{\operatorname{cn}^2 u \operatorname{dn}^2 u} \text{ dynes per sq. cm.,} \quad . . . (54)$$

when the value of σ given by (30) is used. The total force per cm. depth acting normally on any length of any conductor in the system is, therefore, $\int_{z_1}^{z_2} X_1 dz$, where the limits are the values of z at the two ends of the range of surface considered. Call this total force X,

$$X = \int_{z_1}^{z_2} X_1 dz \quad (55)$$

$$= \int_{u_1}^{u_2} \frac{(2E - k'^2 K)^2 \operatorname{sn}^2 u}{2\pi a^2 \operatorname{cn}^2 u \operatorname{dn}^2 u} \cdot \frac{dz}{du} \cdot du, \quad (56)$$

replacing dz by $\frac{dz}{du} \cdot du$.

Insert $\frac{dz}{du} = \frac{a i \operatorname{cn}^2 u \operatorname{dn}^2 u}{(2E - k'^2 K) \operatorname{sn}^2 u}$

from equations (3) and (4).

After integrating it is found that

$$X = \frac{i(2E - k'^2 K)(u_2 - u_1)}{2\pi a} \quad (57)$$

Let $\frac{(2E - k'^2 K)}{2\pi a}$ be called M. (58)

Then $X = Mi(u_2 - u_1)$ (59)

On a length LO, *i. e.* half the length of the central plane in fig. 1, if it is a conductor, the force is

$$X = Mi \left(0 - \frac{iK'}{2} \right) = \frac{MK'}{2} \text{ dynes per cm. depth,} \quad \dots (60)$$

since, at L, $u_2 = 0$. and at O, $u = \frac{iK'}{2}$. On the whole central plane the normal force is

$$MK' = \frac{(2E - k'^2 K) K'}{2\pi a} \text{ dynes per cm. depth.} \quad \dots (61)$$

On any length of the central plane, if it is a conductor, extending from O to any point whose parameter is $u = iq$, the total normal force is

$$\begin{aligned} X &= Mi \left(u - \frac{iK'}{2} \right) \\ &= M \left(\frac{K'}{2} - q \right). \quad \dots (62) \end{aligned}$$

The co-ordinate x of the point in question is obtained by inserting the value of q in equation (27).

On the flat end of an electrode the normal force is again

$$X = Mi(u_2 - u_1), \text{ by (58).}$$

At a corner D', $u = K$ and at B', $u = K + iK'$, hence

$$\begin{aligned} X &= Mi(K - K - iK') \\ &= MK' \text{ dynes per cm. depth.} \quad \dots (63) \end{aligned}$$

This force acts towards the central plane and agrees, as it should, with the result of (61), expressing the force on the whole central plane. The normal force on one of the side-faces of an electrode is equal to

$$[Mi(u_2 - u_1)] = MK \text{ dynes per cm. depth,} \quad \dots (64)$$

since at D', $u = K$ and at E', $u = 0$.

In the magnetic system of pole-pieces with a gap of the form shown in fig. 1. the mechanical force on a small magnetic body placed at any point in the gap is proportional to $\frac{dF^2}{dz}$, where F is defined by equation (28)

$$i. e., F = \left| \frac{dW}{dz} \right|.$$

Hence, using equation (10), squaring and differentiating, it is found that $\frac{dF^2}{dz}$ is proportional to $1 - k^2 \operatorname{sn}^4 u$, to $\operatorname{sn}^3 u$, and inversely proportional to $\operatorname{cn}^5 u$ and $\operatorname{dn}^5 u$.

Hence the force is zero when $\operatorname{sn} u = \pm \frac{1}{\sqrt{k}}$ and $\operatorname{sn} u = 0$. The first result corresponds to points in the middle of the ends of the pole-pieces, where $u = \pm \left(K + \frac{iK'}{2} \right)$ and the second to points at infinity, where $u = 0$. The force is infinite at the corners where $\operatorname{cn} u = 0$ and $\operatorname{dn} u = 0$.

§ 11. *Experimental Test of the Theory.*

As is well known, the foregoing type of theoretical work applies to corresponding systems in at least six branches of pure or applied physics, viz.: (1) electrostatics, (2) magnetism, (3) hydrodynamics, (4) current electricity, (5) heat flow, and (6) diffusion of solutes in liquids. The assertion that the lines in fig. 7 are equipotentials of certain values has been tested experimentally by a modernized application of an old method involving current electricity (Adams, Proc. Roy. Soc. xxiv, 1875-6, p. 1). To represent the electrically resistive medium a flat sheet of flannelette or blotting-paper soaked in copper sulphate solution was employed. As approximately perfectly conducting electrodes, pieces of sheet brass provided with terminals were used. One was a long rectangular strip, cut to the width of the upper electrode in fig. 7. This was placed opposite another long electrode with a long straight side, representing the central plane of fig. 7. The pair represented the upper half of fig. 7 to size. As a source of current was used a beat frequency oscillator, set for a frequency of 1,000 cycles per second and giving an output of 10 volts r.m.s., was applied across the electrodes. Equipotentials were found by using telephones or A.C. voltmeter. Wires leading to the telephones were fitted to the ends with pointed steel probes. When, on pressing the two point probes into the flannelette, no sound could be heard, it was assumed that the points were on the same equipotential curve. To register the points a sheet of white paper with carbon paper above it was placed underneath the flannelette, and the probes were pressed down a little harder so as to register a dot properly. When, instead of the telephones a voltmeter was used, and the probe at one end was connected to the electrode representing the central plane at zero volts potential and the other probe was moved about until the voltmeter registered 3.30 volts, it was assumed that a point on the nearest equipotential to the central plane in fig. 7 had been found. Other points on the same equipotential were found. The procedure was repeated to get equipotentials at 5.41, 7.36 and 8.63 volts, respectively. Smooth curves drawn through the dots on the white paper underneath were in almost perfect agreement with those in fig. 7. The only slight difference was that at the right and left ends farthest from the gap, the experimental curves curled slightly back towards the central plane, due probably to the fact that in the theoretical case the central plane was infinite in length, whereas the experimental electrode representing it was finite in length, *i. e.*, about 40 cm. long.

§ 12. *Other Special Cases.*

Electrodes of Zero Thickness.—When $b=0$ but a remains finite, we get the case when the electrodes are reduced to plates of negligible thickness. The transformation for this case is given in such text-books as Jeans' 'Electricity and Magnetism.'

It is interesting to consider what happens from the present point of view. In equation (4), which is quite general, write $b=0$. This gives

$$k'^2 = \frac{2(K' - E')}{K'},$$

since the denominator of the right-hand side of (4) cannot be infinite. The only value of k' fulfilling this condition is $k'=0$, as can be seen by examination of the Smithsonian Tables. Hence $k=1$. In this case

$$u = \int_0^\phi \frac{d\phi}{\sqrt{1-k^2 \sin^2 \phi}} \text{ reduces to } \int_0^\phi \frac{d\phi}{\sqrt{1-\sin^2 \phi}} = \frac{1}{2} \log \left(\frac{1+\sin \phi}{1-\sin \phi} \right),$$

and $\text{sn } u$ degenerates, becoming equal to $\tanh u$.

$$\sin \phi = \text{sn } u = \tanh u. \quad (65)$$

We also get

$$z = -a \sinh W \quad (66)$$

and

$$z = -ai \coth 2u \quad (67)$$

The surface density at any point

$$\sigma = \left| \frac{-i}{2a\pi} \sinh 2u \right|. \quad (68)$$

Since

$$\coth u = \frac{\cosh u}{\sinh u} = \frac{\sqrt{1+\sinh^2 u}}{\sinh u},$$

we soon find that

$$\sinh u = \sqrt{\frac{1}{\coth^2 u - 1}}$$

and

$$\sinh 2u = \sqrt{\frac{1}{\coth^2 2u - 1}} \quad (69)$$

By (67),

$$\coth 2u = \frac{-z}{ai}.$$

Hence

$$\begin{aligned} \sigma &= \left| \frac{-i}{2a\pi} \sinh 2u \right|, \text{ by (68),} \\ &= \left| \frac{-i}{2\pi a} \sqrt{\frac{1}{\frac{z^2}{(ai)^2} - 1}} \right| \\ &= \frac{1}{2\pi(z^2 + a^2)^{\frac{1}{2}}}. \end{aligned} \quad (70)$$

When $z = +ia$, $\sigma = \infty$.

When $z = i\infty$, $\sigma = 0$, as we expect.

Since (70) applies to any point on any conductor in this system, we have at the middle of the central plane, if a conductor

$$z=0, \sigma = \frac{1}{2\pi a} \quad (71)$$

When the electrodes lose their thickness and become "thin," the total charge per cm. depth on any portion of one of them

$$=Q_1 = \left| \frac{1}{2\pi} \log \frac{\text{sn } u_2}{\text{sn } u_1} \right|, \text{ by (41),}$$

reduces to
$$Q_1 = \left| \frac{1}{2\pi} \log \frac{\tanh u_2}{\tanh u_1} \right| \dots \dots \dots (72)$$

which reduces to
$$Q_1 = \frac{1}{2\pi} \log \left\{ \frac{z_2 \pm \sqrt{z_2^2 + a^2}}{z_1 \pm \sqrt{z_1^2 + a^2}} \right\} \dots \dots \dots (73)$$

when we write $\coth 2u = -\frac{z}{a^2}$, by (67),

and
$$\coth 2u = \frac{1 + \tanh^2 u}{2 \tanh u}$$

by the properties of hyperbolic functions, and then solve for $\tanh u$ in terms of z .

Similarly, the capacity between any equal and opposite portions of the two electrodes, which for thick electrodes

$$= C_1 = \frac{1}{4\pi^2} \log (k \operatorname{sn}^2 u_2) \text{ e.s.u. per cm. depth, by (44),}$$

becomes $C_1 = \frac{1}{4\pi^2} \log (\tanh^2 u_2)$, here

$$= \frac{1}{2\pi^2} \log \tanh u_2. \dots \dots \dots (74)$$

The expression

$$C_1 = \frac{1}{4\pi^2} \log k$$

given by (49), for the capacity between opposite end faces only, reduces to

$$C_1 = \frac{1}{4\pi^2} \log 1 = 0, \dots \dots \dots (75)$$

as expected when $k=1$ and the thickness becomes zero.

The electric resistance between any equal and opposite portions of the two electrodes when the gap is filled with a resistive medium is

$$R_1 = \frac{\pi \rho}{\log (k \operatorname{sn}^2 u_2)},$$

by (52), for thick electrodes. This reduces to

$$R_1 = \frac{\pi \rho}{\log \tanh^2 u_2} = \frac{\pi \rho}{2 \log \tanh u_2} \text{ e.m.u. per cm. depth} \dots \dots (76)$$

for thin electrodes.

TABLE X. (See fig. 7.)

Curve No.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
H ₂	0	-0.025	-0.05	-0.1	-0.2	∞	+0.2	+0.1	0.05	0.025	0
Volts	+10	+8.63	+7.36	+5.41	+3.30	0	-3.30	-5.41	-7.36	-8.63	-10

XCVI. *Alternating Loads on Sleeve Bearings.*

By J. DICK, B.Sc., Ph.D.*

[Received January 17, 1944.]

THE hydrodynamic theory of lubrication has been used in studying many conditions in complete sleeve bearings, but no solution has been found giving the attitude-eccentricity conditions in the case of a bearing subjected to a simple sinusoidal load. Some cases of varying loads have been studied by Harrison⁽¹⁾, Robertson⁽²⁾, and Swift⁽³⁾, and a common deduction in all three publications is that of the free locus of the centre of a shaft when the shaft is rotating with no load and the shaft axis is displaced from the bearing axis. This locus is shown to be a circle, and the angular velocity of "whirl" is one half of the angular velocity of the shaft. Harrison and Swift also studied the free locus for the case in which a steady load was applied, and the shaft was displaced from the position of equilibrium. In this case, if damping factors are neglected, the locus is a closed path, and the time for a complete cycle depends on the magnitude of the steady load and the extent of the initial displacement. If the initial displacement is large enough to give a locus enclosing most of the clearance circle, the time to describe the complete cycle is approximately the same as that for the no-load locus.

The case of a small oscillating load superposed on a relatively large steady load was also studied by Harrison, but the solution is valid for only a very limited range of cases. The behaviour of the film under certain cyclic loads with zero mean load was studied by Swift. In this case the load at any instant was a function of the angular displacement of the line joining the shaft and bearing centres, the function chosen being one that permitted integration to be carried out. The load variation departed considerably from a sinusoidal variation, except when the locus was almost circular, in which case the conditions approached those of the free locus with no load. This suggested that with sinusoidal loading a critical condition could arise when the frequency of the load was in the region of $\frac{1}{2} \frac{\omega_0}{2\pi}$, where ω_0 is the angular velocity of the journal.

The present article considers the case where the locus of the shaft centre is an ellipse, and the shaft centre is assumed to move round the ellipse in such a way that the hydrodynamic load along one axis of the ellipse is zero. The hydrodynamic load along the other axis is then compared with a sinusoidal load.

* Communicated by the Author.

Fig. 1 shows the conventional diagram indicating the position of the shaft centre relative to the bearing centre. The symbols used for this diagram, and for the analysis which follows, are :

R = radius of journal.

R' = radius of bearing.

r = radial clearance = $R' - R$.

λ = coefficient of viscosity.

OJ = eccentricity of shaft centre, O being the bearing centre and J the shaft centre.

$\epsilon = \frac{OJ}{r}$, giving eccentricity as a dimensionless quantity.

F = resultant hydrodynamic load on the bearing at a given instant, per unit length of bearing.

$P = \frac{F}{12\pi\lambda R} \left(\frac{r}{R}\right)^2$ is taken as the "load criterion," per unit length of bearing.

Fig. 1.

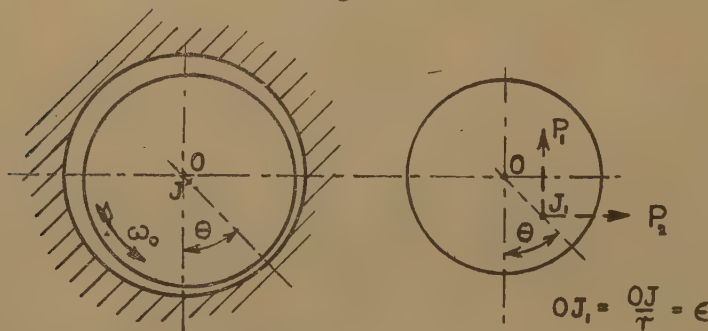


Diagram showing attitude θ and eccentricity ϵ .

The position of the shaft centre is defined by the polar co-ordinates ϵ and θ , and in this diagram, since ϵ cannot be greater than unity, the locus must lie inside a circle of unit radius.

Further symbols used are :

ω_0 = angular velocity of journal about J .

$d\theta/dt$ = angular velocity of OJ .

T = period of one complete cycle.

$p = 2\pi/\tau$ = average angular velocity of OJ .

The locus to be considered is an ellipse, of semi-axes ϵ_1 and ϵ_2 , with $\epsilon = \epsilon_1$ when $\theta = 0$. It will also be assumed that the hydrodynamic load in the direction $\theta = \frac{\pi}{2}$ is zero. The load in the direction $\theta = 0$ will then be compared with $P_0 \sin pt$.

In the comparison, P_0 is defined as the load which, applied statically, gives an eccentricity ϵ_0 .

The work of Harrison, Robertson, and Swift showed that the load components P_1 and P_2 (fig. 1) are given by

$$P_1 = \frac{F_1}{12\pi\lambda R} \left(\frac{r}{R} \right)^2 = \frac{\epsilon \sin \theta}{(2+\epsilon^2)\sqrt{1-\epsilon^2}} \left(\omega_0 - 2 \frac{d\theta}{dt} \right) + \frac{\cos \theta}{(1-\epsilon^2)^{3/2}} \frac{d\epsilon}{dt}; \quad (1)$$

$$P_2 = \frac{F_2}{12\pi\lambda R} \left(\frac{r}{R} \right)^2 = \frac{\epsilon \cos \theta}{(2+\epsilon^2)\sqrt{1-\epsilon^2}} \left(\omega_0 - 2 \frac{d\theta}{dt} \right) - \frac{\sin \theta}{(1-\epsilon^2)^{3/2}} \frac{d\epsilon}{dt}. \quad (2)$$

The conditions to be considered in this article are (a) an elliptic locus and (b) $P_2=0$.

The equation of the ellipse is

$$\epsilon^2 = \frac{2\epsilon_1^2\epsilon_2^2}{(\epsilon_1^2+\epsilon_2^2) - (\epsilon_1^2-\epsilon_2^2)\cos 2\theta}, \quad \dots \quad (3)$$

from which

$$\frac{d\epsilon}{dt} = -\frac{\epsilon^3}{2\epsilon_1^2\epsilon_2^2} (\epsilon_1^2-\epsilon_2^2) \sin 2\theta \frac{d\theta}{dt}. \quad \dots \quad (4)$$

Putting $P_2=0$, we have

$$\left(\omega_0 - 2 \frac{d\theta}{dt} \right) = \frac{2+\epsilon^2}{\epsilon(1-\epsilon^2)} \tan \theta \frac{d\epsilon}{dt}. \quad \dots \quad (5)$$

From (3), (4), and (5)

$$\frac{\omega_0}{2\epsilon_2^2} = \left\{ \frac{3(1-\epsilon_1^2)}{(\epsilon_1^2+\epsilon_2^2-2\epsilon_1^2\epsilon_2^2) - (\epsilon_1^2-\epsilon_2^2)\cos 2\theta} - \frac{1}{(\epsilon_1^2+\epsilon_2^2) - (\epsilon_1^2-\epsilon_2^2)\cos 2\theta} \right\} \frac{d\theta}{dt}. \quad (6)$$

Integration can be effected by using the parameter γ , where

$$(a-b\cos 2\theta)(a+b\cos 2\gamma) = a^2 - b^2.$$

This gives

$$\cos 2\gamma = \frac{a\cos 2\theta - b}{a - b\cos 2\theta} \quad \text{and} \quad \tan \gamma = \sqrt{\frac{a+b}{a-b}} \tan \theta.$$

Also

$$\frac{d\theta}{a-b\cos 2\theta} = \frac{d\gamma}{\sqrt{a^2-b^2}}.$$

The values of γ are equal to those of θ when $\theta=0, \pi/2, \pi, 3\pi/2, 2\pi$, etc.

Hence

$$\begin{aligned} \int_0^\theta \frac{d\theta}{a-b\cos 2\theta} &= \left[\frac{2\gamma}{2\sqrt{a^2-b^2}} \right]_{\theta=0}^{\theta=\theta} \\ &= \frac{1}{2\sqrt{a^2-b^2}} \cos^{-1} \left(\frac{a\cos 2\theta - b}{a - b\cos 2\theta} \right). \end{aligned}$$

Also

$$\int_0^{2\pi} \frac{d\theta}{a-b\cos 2\theta} = \left[\frac{\gamma}{\sqrt{a^2-b^2}} \right]_0^{2\pi} = \frac{2\pi}{\sqrt{a^2-b^2}}.$$

Assuming $t=0$, when $\theta=0$, these integrations, applied to (6), give

$$\begin{aligned} t = \frac{1}{2\omega_0} \cdot \frac{\epsilon_2}{\epsilon_1} &\left[3\sqrt{\frac{1-\epsilon_1^2}{1-\epsilon_2^2}} \cos^{-1} \left\{ \frac{(\epsilon_1^2+\epsilon_2^2-2\epsilon_1^2\epsilon_2^2)\cos 2\theta - (\epsilon_1^2-\epsilon_2^2)}{(\epsilon_1^2+\epsilon_2^2-2\epsilon_1^2\epsilon_2^2) - (\epsilon_1^2-\epsilon_2^2)\cos 2\theta} \right\} \right. \\ &\quad \left. - \cos^{-1} \left\{ \frac{(\epsilon_1^2+\epsilon_2^2)\cos 2\theta - (\epsilon_1^2-\epsilon_2^2)}{(\epsilon_1^2+\epsilon_2^2) - (\epsilon_1^2-\epsilon_2^2)\cos 2\theta} \right\} \right], \quad \dots \quad (7) \end{aligned}$$

and the period T for a complete cycle is

$$T = \frac{2\pi}{\omega_0 \epsilon_1} \left\{ 3 \sqrt{\frac{1-\epsilon_1^2}{1-\epsilon_2^2}} - 1 \right\}. \quad (8)$$

The component P_1 is found by substituting for $\frac{d\epsilon}{dt}$ and $\frac{d\theta}{dt}$ from equations (4) and (6) in equation (1), and on simplifying this gives

$$P_1 = - \frac{\omega_0 \epsilon \sin \theta}{(1-\epsilon^2)^{1/2} \left\{ (1-\epsilon^2) \left(\frac{2\epsilon_2^2}{\epsilon_1^2 - \epsilon_2^2} \right) - 3\epsilon^2 \sin^2 \theta \right\}}. \quad (9)$$

The object now is to compare this with a sinusoidal load, *i. e.*, to compare it with

$$P_1 = P_0 \sin \frac{2\pi}{T} t. \quad (10)$$

Approximate Solution for Small Values of P_0 .

Calculated results for a moderate value of P_0 will be considered later in this article, but since the expressions (7) and (9) do not, in their present form, give any indication of the probable degree of approximation, it is proposed to consider in the first instance the case in which ϵ^2 , ϵ_1^2 , ϵ_2^2 , and $3\epsilon^2 \sin^2 \theta$ are small compared with unity. The equation (7) becomes

$$t = \frac{1}{\omega_0 \epsilon_1} \frac{\epsilon_2}{\cos^{-1} \left(\frac{(\epsilon_1^2 + \epsilon_2^2) \cos 2\theta - (\epsilon_1^2 - \epsilon_2^2)}{(\epsilon_1^2 + \epsilon_2^2) - (\epsilon_1^2 - \epsilon_2^2) \cos 2\theta} \right)}. \quad (7a)$$

and equation (9) becomes

$$P_1 = \frac{\omega_0}{2} \left(\frac{\epsilon_2^2 - \epsilon_1^2}{\epsilon_2^2} \right) \epsilon \sin \theta. \quad (9a)$$

Let
$$\cos 2\gamma = \frac{(\epsilon_1^2 + \epsilon_2^2) \cos 2\theta - (\epsilon_1^2 - \epsilon_2^2)}{(\epsilon_1^2 + \epsilon_2^2) - (\epsilon_1^2 - \epsilon_2^2) \cos 2\theta}.$$

This gives
$$(1 - \cos 2\gamma) = \frac{\epsilon^2}{\epsilon_2^2} (1 - \cos 2\theta);$$

hence
$$\sin \gamma = \frac{\epsilon}{\epsilon_2} \sin \theta.$$

Also, from (7a)
$$\gamma = \frac{\omega_0}{2} \frac{\epsilon_1}{\epsilon_2} t,$$

and from (9a)
$$\begin{aligned} P_1 &= \frac{\omega_0}{2} \left(\frac{\epsilon_2^2 - \epsilon_1^2}{\epsilon_2^2} \right) \epsilon_2 \sin \gamma \\ &= \frac{\omega_0}{2} \epsilon_2 \left\{ 1 - \left(\frac{\epsilon_1}{\epsilon_2} \right)^2 \right\} \sin \frac{\omega_0}{2} \frac{\epsilon_1}{\epsilon_2} t, \end{aligned} \quad (11)$$

which is a sinusoidal load.

The value of P_0 given by equation (11) is

$$P_0 = \frac{\omega_0}{2} \epsilon_2 \left\{ 1 - \left(\frac{\epsilon_1}{\epsilon_2} \right)^2 \right\}.$$

If a steady load P_0 gives eccentricity ϵ_0 , which is small, then

$$\epsilon_0 = \epsilon_2 \left\{ 1 - \left(\frac{\epsilon_1}{\epsilon_2} \right)^2 \right\} \text{ approximately.}$$

That is,
$$\epsilon_2 = \frac{\epsilon_0}{1 - \left(\frac{\epsilon_1}{\epsilon_2} \right)^2} \quad \dots \dots \dots (12)$$

From this equation it can be seen that when ϵ_1 and ϵ_2 are nearly equal, that is, when the frequency is approximately $\frac{1}{2} \frac{\omega_0}{2\pi}$, the values of ϵ_1 and ϵ_2 are both large. The approximation used in deriving expression (11) requires the eccentricity to be small, hence this expression is not valid when the frequency is near $\frac{1}{2} \frac{\omega_0}{2\pi}$, even if ϵ_0 is small. However, when $\epsilon_1^2 - \epsilon_2^2$ is small compared with ϵ_1^2 and ϵ_2^2 , the approximations from (7) and (9) become

$$t = \frac{\epsilon_2}{\epsilon_1} \frac{2\theta}{\omega_0} \quad \dots \dots \dots (7b)$$

and
$$P = \frac{\omega_0}{2} \left(\frac{\epsilon_1^2 - \epsilon_2^2}{\epsilon_2^2} \right) \frac{\epsilon \sin \theta}{(1 - \epsilon^2)^{3/2}} \quad \dots \dots \dots (9b)$$

From these equations

$$P = \frac{\omega_0}{2} \left(\frac{\epsilon_1^2 - \epsilon_2^2}{\epsilon_2^2} \right) \frac{\epsilon}{(1 - \epsilon^2)^{3/2}} \sin \frac{\omega_0}{2} \frac{\epsilon_1}{\epsilon_2} t,$$

or, since $\epsilon = \epsilon_1 = \epsilon_2$ approximately,

$$P = \frac{\omega_0}{2} \left(\frac{\epsilon_1^2 - \epsilon_2^2}{\epsilon_2^2} \right) \frac{\epsilon_2}{(1 - \epsilon_2^2)^{3/2}} \sin \frac{\omega_0}{2} \frac{\epsilon_1}{\epsilon_2} t. \quad \dots \dots \dots (13)$$

This shows that when the frequency is near $\frac{1}{2} \frac{\omega_0}{2\pi}$, the elliptical locus gives a load that is very nearly sinusoidal; hence it can be said that a sinusoidal load of small amplitude will give a locus which is a close approximation to an ellipse for any value of $\frac{2\pi}{\omega_0 T}$.

It will be noted that the expression for P_0 given by equation (12) differs from that in equation (11), the difference being due to the factor $\frac{1}{(1 - \epsilon_2^2)^{3/2}}$, which would be approximately unity for the conditions under which (11) is valid, that is, when ϵ is small.

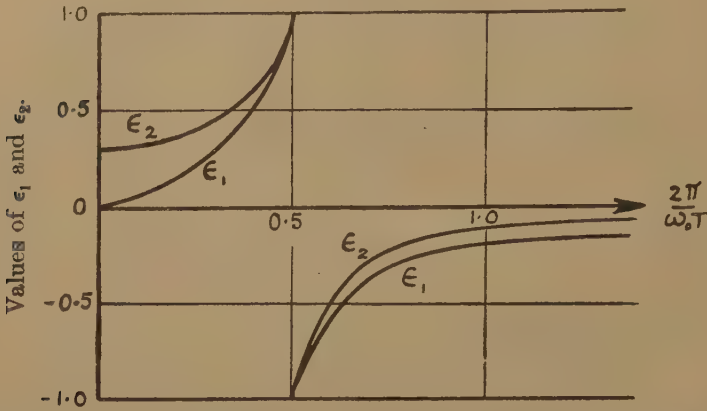
Check for Larger Load.

A study of the terms neglected in deducing (7a) and (9a) suggests $\epsilon_0 = 0.3$ as a suitable value for comparing P_1 of equation (9) with P_1 of equation (10). The approximation is likely to improve rapidly as ϵ_0 is decreased below this value, and to become rapidly worse as ϵ_0 is increased above this value.

When $\epsilon_0 = 0.3$,
$$P_0 = \frac{\omega_0 \epsilon_0}{(2 + \epsilon_0^2) \sqrt{1 - \epsilon_0^2}} = 0.1505 \omega_0.$$

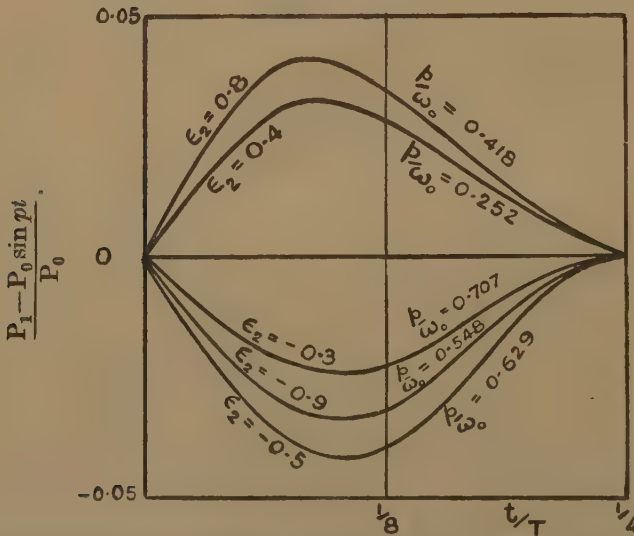
Taking $P_1 = 0.1505 \omega_0$ when $\theta = \frac{\pi}{2}$, values of ϵ_1 and ϵ_2 can be determined for various values of $\frac{2\pi}{\omega_0 T}$. Fig. 2 shows these in graphical form.

Fig. 2.



Semi-axes ϵ_1 and ϵ_2 of ellipses for which $\epsilon_0 = 0.3$ and $P_0 = \frac{\epsilon_0}{(2 + \epsilon_0^2) \sqrt{1 - \epsilon_0^2}}$.

Fig. 3.



Differences between loads for elliptic paths and sinusoidal load given as fraction of maximum load.

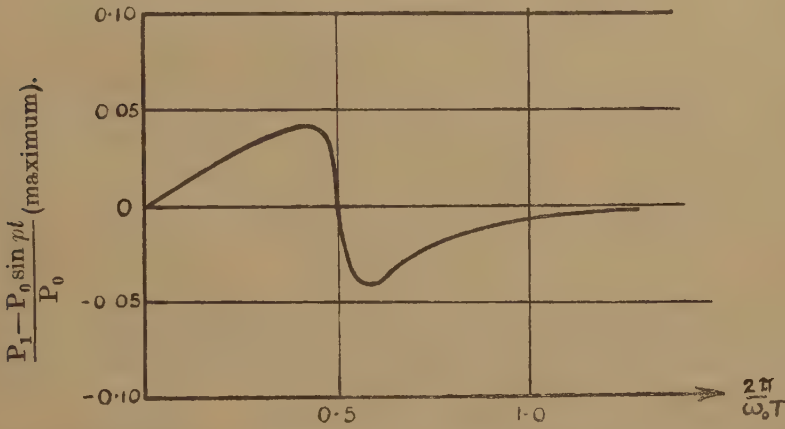
The error we are considering can be expressed in the form

$$\frac{P_1 - P_0 \sin \frac{2\pi}{T} \epsilon}{P_0}.$$

This is calculated by choosing a value of θ with a particular ellipse, then calculating t from equation (7) and P_1 from equation (9). Fig. 3 shows the calculated errors for several ellipses over one quarter cycle, and fig. 4 shows a curve of maximum errors for all ellipses.

The greatest error for this particular value of P_0 is about 4 per cent. of the maximum load; hence it can be said that the ellipse gives a good approximation to the locus for a sinusoidal load of moderate amplitude. A few calculations with large amplitudes showed much larger errors; but it would appear that the elliptical locus would give a satisfactory first approximation for a considerable range of cases.

Fig. 4.



Curve showing maximum value of $\frac{P_1 - P_0 \sin pt}{P_0}$ for various ellipses for

$$\text{which } \epsilon_0 = 0.3 \text{ and } P_0 = \frac{\epsilon_0}{(2 + \epsilon_0^3) \sqrt{1 - \epsilon^2}}.$$

General Conclusions.

It is obvious that the elliptical locus gives a load near enough to a sinusoidal load to justify using the ellipse for descriptive purposes, and the following general conclusions can be drawn for the case of a sinusoidal load $P_0 \sin pt$, where P_0 is defined as the load giving an eccentricity ϵ_0 when applied statically.

When p/ω_0 is small, the locus is a line at right angles to the load. J_1 moves along this line between the limits $\pm \epsilon_0$.

With increase of frequency between 0 and $\frac{1}{2} \frac{\omega_0}{2\pi}$, the locus is approximately elliptical, with ϵ_2 greater than ϵ_0 and ϵ_1 small at first; but as the frequency increases, ϵ_2 , ϵ_1 and the ratio ϵ_1/ϵ_2 all increase.

As the frequency approaches the value $\frac{1}{2} \frac{\omega_0}{2\pi}$, that is, when p/ω_0 approaches $\frac{1}{2}$, ϵ_1 and ϵ_2 are nearly equal and both are large. When

$p/\omega_0 = \frac{1}{2}$, a critical condition is obtained, in which $\epsilon_1 = \epsilon_2$, and the film is not capable of maintaining the sinusoidal load.

At higher frequencies the locus is again approximately elliptical, but in this case ϵ_2 is less than ϵ_1 , and with increase in frequency, ϵ_1 , ϵ_2 and the ratio ϵ_2/ϵ_1 all decrease.

At very high frequencies the conditions approach those of the non-rotating shaft to which a sinusoidal load is applied. In this case the range for symmetrical oscillations with $\epsilon_2 = 0$ has been shown by Swift to be $\pm \frac{\Delta_0}{\sqrt{1-\Delta_0^2}}$, where $\Delta_0 = \frac{P_0 r^2}{12\lambda \pi p R^3}$ for a load $P_0 \sin pt$ per unit length of bearing.

As the radius vector OJ traces out the path it does not move with uniform angular velocity. The angular velocity is $\frac{\omega_0}{2}$ when $\theta = 0$ and $\theta = 180^\circ$, and at other angles it is given approximately by the expression $\frac{d\theta}{dt} = \frac{\omega_0}{2} + \omega_0 a(1 - \cos 2\theta)$. When p/ω_0 is small $a = -\frac{1}{4}$, and a increases with increase in p/ω_0 , being zero when $p/\omega_0 = \frac{1}{2}$ and positive for higher values of p/ω_0 .

References.

- (1) Trans. Camb. Phil. Soc. vol. xxii. p. 39 (1912-23).
- (2) Phil. Mag. vol. xv. p. 113 (1933).
- (3) Journ. Inst. Civ. Engrs. no. 4, p. 161 (1936-37).

XCVII. *An Application of the Method of Finite Difference Equations to a Problem of Bending Moments. (Continuous Beam of N Equal Spans under Uniform Loading.)*

By N. J. DURANT*.

[Received May 25, 1944.]

Summary.

The bending moments at the supports of a continuous beam of N equal spans (where N is any positive integer) are obtained explicitly from a formula which is the solution of a Finite Difference Equation. A simple example shows the saving in time which is thus effected, and exhibits the power of the method of Finite Difference Equations in solving an otherwise cumbersome problem.

1. The bending moments in a continuous beam are usually derived from the solution of a number of simultaneous linear equations, or by a method of moment distribution, either of which entails a considerable amount of

* Communicated by the Author.

labour even for small values of N . We propose the following alternative approach.

Consider a beam of N spans each of length L , continuous over $N+1$ supports on the same level, and continuously loaded with an intensity of load of w per unit length.

Suppose the origin to be at the extreme left support; and let n be the number of any support to the right of the origin, so that n may take the values $0, 1, 2 \dots, N$, there being $N+1$ supports altogether.

The bending moments at any three successive supports are connected by the theorem of three moments, thus :

$$M_{n-1} + 4M_n + M_{n+1} = -\frac{1}{2}wL^2 \quad (n=1, 2 \dots, N-1). \quad (1)$$

Equation (1) is, however, a second order finite difference equation.

To reduce it to a recurrence formula, substitute

$$u_n = M_n + \frac{1}{12}wL^2 \quad (n=0, 1, 2, \dots, N). \quad (2)$$

Equation (1) then becomes :

$$u_{n-1} + 4u_n + u_{n+1} = 0. \quad (3)$$

Now make the standard substitution $u_n = E^n$; then, from (3) . . . (4)

$$E^{n-1}(1+4E+E^2)=0. \quad (5)$$

Hence $E = -2 \pm \sqrt{3} \equiv E_1$ and E_2 respectively, . . . (6)

the value of $E=0$ being of no significance, since zero roots contribute nothing to the results.

$$E_1 = -0.268, \quad (7)$$

$$E_2 = -3.732. \quad (8)$$

The general solution of equation (3) is

$$u_n = AE_1^n + BE_2^n. \quad (9)$$

But

$$E_1 E_2 = 1. \quad (10)$$

$$\therefore u_n = AE_1^n + BE_1^{-n}, \quad (11)$$

and the general solution of equation (1) is

$$M_n = AE_1^n + BE_1^{-n} - \frac{1}{12}wL^2. \quad (12)$$

The parameters A and B are determined from the boundary conditions, which in this case are

$$M_n = 0 \text{ where } n=0 \text{ and where } n=N.$$

From these conditions we find that

$$A = \frac{1}{12}wL^2 \frac{E^{-N}-1}{E^{-N}-E^N}, \quad (13)$$

$$B = \frac{1}{12}wL^2 \frac{1-E^N}{E^{-N}-E^N}. \quad (14)$$

$$\therefore M_n = \frac{1}{12}wL^2 \left\{ \frac{E^{-(N-n)} - E^n + E^{-n} - E^{(N-n)}}{E^{-N} - E^N} - 1 \right\}. \quad (15)$$

2. Formula (15) can be put into a more convenient form if we write $E = -e^{-\gamma} = -0.268$ (where e is the base of Napierian logarithms).

From this $\gamma = 1.317$ (16)

$$\therefore M_n = \frac{1}{12} wL^2 \left\{ (-1)^n \frac{e^{\gamma(N-n)} - e^{-\gamma(N-n)} + e^{\gamma n} - e^{-\gamma n}}{e^{\gamma N} - e^{-\gamma N}} - 1 \right\}; \quad (17)$$

$$i. e., \quad M_n = \frac{1}{12} wL^2 \left\{ (-1)^n (\cosh \gamma n - \tanh \frac{1}{2} \gamma N \sinh \gamma n) - 1 \right\}. \quad (18)$$

From this formula the bending moment and its sign at any support of a continuous beam can be calculated immediately without the necessity of calculating any moment not actually required.

As the beam is assumed to be symmetrical and symmetrically loaded about the centre of its length, the largest value that need be taken for n is either $\frac{1}{2}N$ or $\frac{1}{2}(N-1)$, according as N is even or odd.

3. If $N \geq 7 \tanh \frac{1}{2} \gamma N = 1$ very nearly, so that formula (18) reduces to the approximate form,

$$M_n \sim \frac{1}{12} wL^2 \{ (-1)^n (\cosh \gamma n - \sinh \gamma n) - 1 \}. \quad (19)$$

Again, from formula (19), for $4 \leq n \leq N-3$

$$M_n = -\frac{1}{12} wL^2. \quad (20)$$

For $0 < n < 4$ and for $N-3 < n < N+1$ we therefore have

$$M_1 = M_N = \frac{1}{12} wL^2 \{ -\cosh \gamma + \sinh \gamma - 1 \} = \frac{1}{12} wL^2 (-1.268), \quad (21)$$

$$M_2 = M_{N-1} = \frac{1}{12} wL^2 \{ \cosh 2\gamma - \sinh 2\gamma - 1 \} = \frac{1}{12} wL^2 (-0.928), \quad (22)$$

$$M_3 = M_{N-2} = \frac{1}{12} wL^2 \{ -\cosh 3\gamma + \sinh 3\gamma - 1 \} = \frac{1}{12} wL^2 (-1.019). \quad (23)$$

Hence, for beams of seven spans or more (*i. e.* $N \geq 7$), the moments at the internal supports are given by formula (20), and those at the supports near each end are given by formulæ (21), (22) and (23).

4. It should be observed that the formulæ give the bending moments at the supports and not at intermediate points along the spans. However, when the bending moments at the supports are known the problem of determining the bending moments at intermediate points is one of simple statics.

Moreover, when $N \geq 7$ the positions of maximal and minimal bending moments can be determined once for all from formulæ (20), (21), (22) and (23).

I am indebted to Dr. J. Bronowski and Dr. F. Garwood for valuable criticism.

XCVIII. "*Transient Response in Frequency Modulation,*"
by Mr. D. A. Bell in the *Phil. Mag.* p. 143 (March, 1944).

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

REFERRING to the March, 1944, issue of the 'Philosophical Magazine and Journal of Science,' particularly to the article on "Transient Response in Frequency Modulation," by D. A. Bell, I would like to bring the following points to your notice:—

Page 145, equation (2)

$$\bar{\omega} = \frac{d}{dt} \{ \omega t + \psi \} = \dots$$

should obviously read

$$\bar{\omega} = \frac{d}{dt} \{ \omega t + \phi \} = \dots$$

Equation (3)

$$\omega = \frac{\omega_1 R^2 + \omega_2 S^2 + (\omega_1 + \omega_2) RS \cos (\omega_1 - \omega_2) t}{R^2 + S^2 + 2RS \cos (\omega_1 - \omega_2) t}$$

should read $\bar{\omega} = \text{etc.}$

On page 146, equation (5)

$$x^2 + p^3$$

should obviously read

$$x^2 + p^2.$$

On page 147, paragraph 2, line 3, it is correct that the numerator of each reduces to

$$(x^2/\omega^2)(1 - x^2/\omega^2 - 3/4 Q^2),$$

but the denominators are different.

Equation 11, same page,

ω should read $\bar{\omega}$.

Page 148, line 29,

$$\omega_2 - \omega_3$$

should obviously read

$$\omega_1 - \omega_2.$$

Another point arising out of this line is: does "finite" exclude zero?

Page 155. I am not clear whether or not lines 1 and 2 do follow necessarily from the previous argument.

Yours faithfully,
P. K. CHATTERJEA.

134 Fellows Road, Swiss Cottage,
London, N.W.3.
October 9th, 1944.

SER. 7, VOL. 35, NO. 251—DEC. 1944.

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

I AM indebted to Mr. P. K. Chatterjea for pointing out a number of errors in the printed symbols in my paper on "Transient Response in Frequency Modulation."

With regard to p. 147, paragraph 2, the condition for $A=B$ should have been stated as governed by the same condition as $C=D$, namely, $1/4Q^2$ small compared with unity.

The argument in parentheses from the lower part of p. 154 to the top of p. 155 might be clearer in this form :—

(1) If the amplitude/frequency characteristic of a receiver with limiter is explored by means of a simple sinusoidal signal of variable frequency, then it is found that over a considerable frequency-band the amplitude of the output is the same for all frequencies.

(2) But if a frequency-modulated signal is applied to the same receiver with limiter, then the output of the limiter consists of a frequency-modulated wave of constant amplitude, which may be resolved by a Fourier analysis into a number of components of different frequencies which all lie within the flat pass band, but yet have different amplitudes.

Yours faithfully,

D. A. BELL.

The Tower, Quaker's Walk,
Winchmore Hill, London, N.21.
October 19th, 1944.

XCIX. "*Transient Response in Frequency Modulation*,"
by Mr D. A. Bell in the *Phil. Mag.* p. 143 (March, 1944).

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

IN the paper by D. A. Bell, "Transient Response in Frequency Modulation" ('*Philosophical Magazine*,' March, 1944), there appears to have been a numerical error in calculation which alters or even vitiates some of the paper's conclusions regarding the effect of passing an F.M. wave through a selective tuned circuit. Equation (24), according to my recalculation, does not lead to the irregularity shown in fig. 9; and hence, fig. 10, which is derived from fig. 9, erroneously shows a peak. Fig. 9 is in error in the range between 30 degrees and 50 degrees; and when this error is corrected the severe distortion in the demodulated F.M. wave is not found.

Respectfully yours,

H. E. CURTIS,

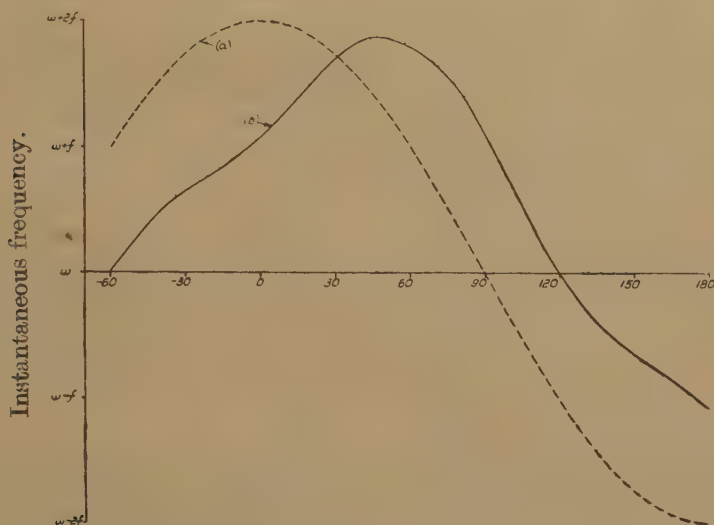
Member of Technical Staff,

Bell Telephone Laboratories,
463 West Street, New York.
October 21st, 1944.

To the Editors of the *Philosophical Magazine*.

GENTLEMEN,—

IN reply to Mr. Curtis, I wish first to apologise to the Editor and to the readers of this Journal for the error in figs. 9 and 10 of my paper. It should have occurred to me that the irregularity shown in fig. 9 was improbable, and of course fig. 10 is entirely governed by fig. 9; the accompanying graph shows a corrected version of the curve given in the original fig. 10, the only substantial difference being the elimination of the sharp peak.



Phase of modulation cycle.

Curve (a): Applied frequency modulation.

Curve (b): Applied modulation distorted by circuit of narrow band-width.

I should also like to express the hope that the Bell Telephone Laboratories may be able to publish some detailed information on the response to frequency-modulated signals of coupled circuits and other complex systems of practical interest.

Yours faithfully,
D. A. BELL.

The Tower, Quaker's Walk,
Winchmore Hill, London, N.21.
December 18th, 1944.

C. Notices respecting New Book.

The Impact and Value of Science. By DOUGLAS W. HILL. [Pp. 88.] (London : Hutchinson's Scientific and Technical Publications, 1944). Price 7s. 6d.

SCIENCE plays so large a part, for good or for ill, in the shaping of civilization, that any sincere and serious attempt to find a way by which its influence can be more widely and beneficially exerted deserves careful consideration. This book is not so much an analysis of the effect of science or an assessment of its value as a plea for a more extensive use of its methods. The author does not believe that there is a widespread appreciation of science, or a desire for the help it can give, except in technical application. He is concerned to show that, beyond this, science has a most important contribution to make, through the application of scientific thought in every sphere of human activity ; and the sections of the book, after the first two on the scientific outlook and on science in industry, deal with science in relation to politics, war, education, religion and leadership. The "science" of the title is not therefore used primarily in the sense of "natural science" or "applied science," but rather in the sense of "scientific method" which is regarded as providing "the power to think logically, dispassionately, impersonally, objectively and thoroughly according to a definite pattern that can be consciously adopted and taught." As such, presumably, science, it is proclaimed, is "more than technology, and is greater than the sciences. It is a system of thought, a philosophy, and a guide to maturity. It is a living thing of beauty and of joy"—and a great deal more besides.

When "science" is used for "scientific thinking," and this in turn is understood in a rather general sense, the view that science should take a more prominent place in many fields hardly requires impassioned advocacy ; for it is almost beyond dispute. It does not follow, however, as the main argument would seem to imply, that the predominant control of affairs should therefore be in the hands of scientists. The author is not completely unaware that the problem is more complex than this. "The scientific method is a necessity of life," he says, "but alone it is not enough. It must be practised by men of good will whose faith in the end they serve is deep-rooted and firm." It appears towards the end of the book that science, or scientific method, in spite of the fanfare with which it was greeted at the beginning, cannot accomplish everything. In the course of the general discussion, much is said that is well worth saying, not only in connection with the general contention that the potential value of the wider application of scientific method is insufficiently recognized, but also on such questions as the form and character of education in science. Unfortunately, there is so much extravagance in statement and confusion in argument that the intrinsic merits of many of the views expressed may not easily gain the recognition that they deserve.

E. C. S.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

INDEX TO VOL. XXXV.

- ACCUMULATION** of chance effects and the Gaussian frequency distribution, 395.
- Aharoni (J.), effect of space charge on electron beams, 36.
- , a general theory of antennæ, 427.
- Allan (G. E.) on the tones of thin cylindrical tubes, 571.
- Alternating loads on sleeve bearings, 841.
- Analysis of hardness, 765.
- Appleton's method of measuring the mutual conductance of a valve, modifications of, 64.
- Application of the micro-absorption factor to problems of lattice distortion and the nature of anti-phase domains, on the, 404.
- Auluck (F. C.) and Kothari (D. S.), a note on the minimum radius for degenerate stellar masses, 783.
- Automatic electric balance for research on aerodynamic stability, the development of an, 588.
- Baber (T. D. H.) and Mirsky (L.), note on certain integrals involving Hermite's polynomials, 532.
- Barrer (R. M.), diffusion in spherical shells, and a new method of measuring the thermal diffusivity constant, 802.
- Bates (L. F.) and Lovering (W. F.), modifications of Appleton's method of measuring the mutual conductance of a valve, 64.
- Bell (D. A.), transient response in frequency modulation, 143.
- Bell (R. P.), eigen-value and eigen-functions for the operator $d^2/dx^2 - |x|$, 582.
- Bloch (A.), a new approach to the dynamics of systems with gyroscopic coupling terms, 315.
- Bloch (A.), on the temperature coefficient of air-cored self-inductances, 693.
- Books, new: J. E. Calthrop, A German Physics Reader, 213; Reports on Progress in Physics, vol. IX (1942-1943), 213; Sir Ambrose Fleming: (1) Mathematics for Engineers, (2) Physics for Engineers, (3) Chemistry for Engineers, (4) Mechanics for Engineers, 351; Prof. P. W. Selwood, Magnetochemistry, 353; Tables of the Bessel Functions $J_0(z)$ and $J_1(z)$ for Complex Arguments, 353; Table of the Reciprocals of the Integers from 100,000 through 200,000, 354; Dr. G. W. Scott-Blair, Survey of General and Applied Rheology, 426; Prof. Max Born, Experiment and Theory in Physics, 426; G. H. Hardy and W. W. Rogosinski, Fourier Series, 498; Sir William C. Dampier, Shorter History of Science, 568; Prof. A. A. Luce, Berkeley's Philosophical Commentaries, 569; Dr. Sandell, Colorimetric Methods for the Determination of Traces of Metals, 641; Tables of Lagrangian Interpolation Coefficients prepared by the Mathematical Tables Project, 713; Table of Circular and Hyperbolic Tangents and Cotangents for Radian Arguments, by the Mathematical Table Project, 714; G. W. Carter, The Simple Calculation of Electrical Transients, 786; D. W. Hill, The Impact and Value of Science, 854.
- Bratt (J. B.) and Wight (K. C.), the development of an automatic electric balance for research on aerodynamic stability, 588.
- Brindley (G. W.) the effect of heterogeneity in powders on the intensity of X-ray reflections, 638.
- Butler (C. C.) and Rymer (T. B.), an electrical circuit for harmonic analysis and other calculations, 606.

- Butler (C. C.) and Rymer (T. B.), a non-recording microphotometer with variable contrast sensitivity, 202.
- Carsten (H. R. F.) and McKerrow (N. W.), the tabulation of some Bessel functions $K_\nu(x)$ and $K_\nu'(x)$ of fractional order, 812.
- Centrifugal method of measuring the surface tensions and interfacial tensions of liquids in capillary tubes, 335.
- Chatterjea (P. K.) on Mr. Bell's paper, "Transient response in frequency modulation," 851.
- Coefficients for numerical integration with central differences, 262.
- Conditions of failure and factors of safety in members subject to alternating stresses, 470.
- Curtis (H. E.) on Mr. Bell's paper "Transient response in frequency modulation," 852.
- Dalzell (D. P.), remarks on electrical units, 188.
- Davy (N.), the field between equal semi-infinite rectangular electrodes or magnetic pole-pieces, 819.
- Dick (J.), alternating loads on sleeve bearings, 841.
- Diffusion in spherical shells, and a new method of measuring the thermal diffusivity constant, 802.
- Dimensions, 420.
- of physical magnitudes, on the (fourth paper), 296.
- — (fifth paper), 616.
- Dingle (H.), a relativistic theory of temperature radiation, 499.
- on the dimensions of physical magnitudes (fourth paper), 296.
- — (fifth paper), 616.
- Duncan (W. J.), some devices for the solution of large sets of simultaneous linear equations, 660.
- Duncanson (W. E.), molecular weight, gram molecule and Avogadro's number, 73.
- , terminology and symbols in physics, 81.
- Durant (N. J.), an application of the method of finite difference equations to a problem of bending moments (continuous beam of N equal spans under uniform loading), 848.
- Dynamics of systems with gyroscopic coupling terms, a new approach to the, 315.
- Effect of heterogeneity in powders on the intensity of X-ray reflections, 638.
- space charge on electron beams, 36.
- Eigen-values and eigen-functions for the operator $d^2/dx^2 - |x|$, 582.
- Electrical circuit for harmonic analysis and other calculations, 606.
- units, remarks on, 188.
- Electron beams, effect of space charge on, 36.
- Energy distribution in the spectrum of a frequency modulated wave.— Part I., 787.
- Equilibrium of a thin slab on an elastic foundation of finite depth, 265.
- "Escalator" process for the solution of Lagrangian frequency equations, 735.
- Evans (D. S.), problems of the transmission of light through tissues and some other media, 300.
- Exercises on a tightrope, 414.
- Expansions in terms of associated Legendre functions, 670.
- Experiments on the six focal lines due to reflexions at the surfaces of two plano-cylindrical lenses, 477.
- Field between equal semi-infinite rectangular electrodes or magnetic pole-pieces, 819.
- Finite difference equations to a problem of bending moments, an application of the method of (continuous beam of N equal spans under uniform loading), 848.
- Fletcher (A.), notes on tables of an integral, 16.
- Fraser (H. D.) and Makinson (K. R.), two methods of measuring ultra high frequency electric fields, 355.
- Friend (J. N.) and Hargreaves (W. D.), viscosities and rheochors of aldehydes, nitriles and of secondary and tertiary amines, 619.
- —, viscosities at the boiling point of some primary amines, cyclohexane and some of its derivatives, 57.

- Friend (J. N.) and Hargreaves (W. D.), viscosities of unsaturated six-membered isocyclic compounds, 136.
- Fürth (R.) and Pringle (R. W.), a new photo-electric method for Fourier synthesis and analysis, 643.
- Further properties of recurrent exponential and probability function pulse wave-forms, 459.
- General theory of antennæ, 427.
- Geometry of demagnetizing curves, and its consequences for permanent magnet materials, 91.
- Gladwin (A. S.), a note on a certain multiple integral, 657.
- , energy distribution in the spectrum of a frequency modulated wave.—Part I., 787.
- Gogate (D. V.) and Upadhyaya (V. N.), thermal transpiration and reversible process for a degenerate gas, 760.
- Goldstein (S.) and Lighthill (M. J.), two-dimensional compressible flow past a solid body in unlimited fluid or symmetrically placed in a channel, 549.
- Graphical computation, 159.
- Green's functions in the theory of heat conduction, note on, 495.
- Green (G.), solutions of problems relating to media in contact by the method of wave-trains, 519.
- , solutions of some problems in viscous flow, 250.
- Grew (K. E.), thermal diffusion in mixtures of molecules of small mass difference, 30.
- Hansel (C. W.), graphical computation, 159.
- Hargreaves (W. D.) and Friend (J. N.), viscosities and rheochors of aldehydes, nitriles and of secondary and tertiary amines, 619.
- , viscosities at the boiling point of some primary amines, cyclohexane and some of its derivatives, 57.
- , viscosities of unsaturated six-membered isocyclic compounds, 136.
- Head (J. W.) and Morris (J.), the "Escalator" process for the solution of Lagrangian frequency equations, 735.
- Heat conduction, on a problem in, 102.
- Helmholtz's electrodynamical potential and the induction coefficients of unclosed currents in quasi-stationary processes, 346.
- Hogg (A. H. A.), equilibrium of a thin slab of an elastic foundation of finite depth, 265.
- Hoselitz (K.), the geometry of demagnetizing curves, and its consequences for permanent magnet materials, 91.
- Houstoun (R. A.), the ultrasonic diffraction grating, 192.
- Influence of air-gases on the first and second positive systems of nitrogen, 600.
- crystal size on the absorption factor as applied to Debye-Scherrer diffraction patterns, 215.
- Integrals involving Hermite's polynomials, note on certain, 532.
- Interferometry, new contributions to, Part I.—Two letters and a reply, 210.
- , Part II.—New interference phenomena with Newton's rings, 120.
- , Part III.—The differential polarization phase change of reflection at a thin silver film, 179.
- , Part IV.—The ghost images and scatter rings of the Fabry-Perot interferometer and their effects in hyperfine structure observations, 229.
- Investigation of the existence of "Magnetic" currents, 342.
- Jaeger (J. C.), some problems involving line sources in conduction of heat, 169.
- Kastner (L. J.) on under-expansion in a steam-nozzle, with reference to an unusual case, 276.
- Kinematic relativity, 241.
- Kothari (D. S.) and Auluck (F. C.), a note on the minimum radius for degenerate stellar masses, 783.
- Lighthill (M. J.) and Goldstein (S.), two-dimensional compressible flow past a solid body in unlimited fluid or symmetrically placed in a channel, 549.

- Lipson (H.) and Rogers (L. E. R.), the measurement of X-ray wavelengths by the powder method: $\text{CrK}\beta_1$ and $\text{MnK}\beta_1$, 544.
- Lovering (W. F.), a push-pull resistance-capacity coupled oscillator, 715.
- on the measurement of the residual parameters of a "Q" meter, 491.
- and Bates (L. F.), modifications of Appleton's method of measuring the mutual conductance of a valve, 64.
- Lowan (A. N.), note on Green's functions in the theory of heat conduction, 495.
- MacDonald (D. K. C.), a note on two definitions of noise figure in radio receivers, 386.
- MacRobert (T. M.), expansions in terms of associated Legendre functions, 670.
- Makinson (K. R.) and Fraser (H. D.), two methods of measuring ultra high frequency electric fields, 355.
- McKerrow (N. W.) and Carsten (H. R. F.), the tabulation of some Bessel functions $K_\nu(x)$ and $K_\nu'(x)$ of fractional order, 812.
- Meyerstein (W.), a centrifugal method of measuring the surface tensions and interfacial tensions of liquids in capillary tubes, 335.
- Millest (D. M.), an investigation of the existence of "Magnetic" currents, 342.
- Minimum radius for degenerate stellar masses; a note on the, 783.
- Mirsky (L.) and Baber (T. D. H.), note on certain integrals involving Hermite's polynomials, 532.
- Molecular weight, gram molecule and Avogadro's number, 73.
- Morris (J.) and Head (J. W.), the "Escalator" process for the solution of Lagrangian frequency equations, 735.
- Multiple integral, a note on a certain, 657.
- Murgatroyd (J. B.), the effect of shape on the thermal endurance of glass rods, 17.
- Neville (E. H.), exercises on a tightrope, 414.
- New algebras and their significance for physics and philosophy, 1.
- photo-electric method for Fourier synthesis and analysis, 643.
- Non-recording microphotometer with variable contrast sensitivity, 202.
- Note on tables of an integral, 16.
- Optimum thickness of powder specimens in X-ray diffraction work, on the, 632.
- Owen (E. A.), note on the solubility of hydrogen in palladium, 50.
- Patankar (V. S.) and Tawde (N. R.), influence of air-gases on the first and second positive systems of nitrogen, 600.
- Peddie (W.), the development of the trichromatic theory of colour vision, 209.
- Pendse (C. G.) on the theory of seismometers, 706.
- Physical realizability of electrical networks having prescribed characteristics, with particular reference to those of the probability function type, 778.
- Pringle (R. W.) and Fürth (R.), a new photo-electric method for Fourier synthesis and analysis, 643.
- Problems of transmission of light through tissues and some other media, 300.
- Properties of liquids at ultra high frequencies, methods for the investigation of, 283.
- Push-pull resistance-capacity coupled oscillator, 715.
- Reception of Ohm's electrical researches by his contemporaries, 371.
- Reformulation of the special theory of relativity, introductory notes on a, 105.
- Relativistic theory of temperature radiation, 499.
- Replies to letters on Mr. Bell's paper "Transient response in frequency modulation," 852, 853.
- Reply to Dr. Brindley's letter on "The effect of heterogeneity in powders on the intensity of X-ray reflections," 640.
- Residual parameters of a "Q" meter, on the measurement of the, 491.

- Restricted problem of the Random Walk, solution of the, 538.
- Rice (S. O.), sums of series of the form $\sum_0 a_n J_{n+\alpha}(z) J_{n+\beta}(z)$, 686.
- Roberts (F. F.) and Simmonds (J. C.), the physical realizability of electrical networks having prescribed characteristics, with particular reference to those of the probability function type, 778.
- , further properties of recurrent exponential and probability function pulse wave-forms, 459.
- Rogers (L. E. R.) and Lipson (H.), the measurement of X-ray wavelength by the powder method: $\text{CrK}\beta_1$ and $\text{MnK}\beta_1$, 544.
- Rymer (T. B.) and Butler (C. C.), an electrical circuit for harmonic analysis and other calculations, 606.
- , a non-recording microphotometer with variable contrast sensitivity, 202.
- Salzer (H. E.), coefficients for numerical integration with central differences, 262.
- Scott-Iversen (P. A.), introductory notes on a reformulation of the special theory of relativity, 105.
- Searle (G. F. C.), experiments on the six focal lines due to reflexions at the surface of two plano-cylindrical lenses, 477.
- Seismometers, on the theory of, 706.
- Serensen (S. V.), conditions of failure and factors of safety in members subject to alternating stresses, 470.
- Silberstein (L.), solution of the restricted problem of the Random Walk, 538.
- , the accumulation of chance effects and the Gaussian frequency distribution, 395.
- Silveira (A. da), on Helmholtz's electro-dynamical potential and the induction coefficients of unclosed currents in quasi-stationary processes, 346.
- Simmonds (J. C.) and Roberts (F. F.), the physical realizability of electrical networks having prescribed characteristics, with particular reference to those of the probability function type, 778.
- Solubility of hydrogen in palladium, note on, 50.
- Solutions of problems relating to media in contact by the method of wave-trains, 519.
- Some devices for the solution of large sets of simultaneous linear equations, 660.
- Some problems involving line sources in conduction of heat, 169.
- Sums of series of the form $\sum_0 a_n J_{n+\alpha}(z) J_{n+\beta}(z)$, 686.
- Tables of an integral, note on, 16
- Tabulation of some Bessel functions $K_\nu(x)$ and $K_\nu'(x)$ of fractional order, 812.
- Tawde (N. R.) and Patankar (V. S.), influence of air-gases on the first and second positive systems of nitrogen, 600.
- Taylor (A.) on the application of the micro-absorption factor to problems of lattice distortion and the nature of anti-phase domains, 404.
- on the optimum thickness of powder specimens in X-ray diffraction work, 632.
- , reply to Dr. Brindley's letter on "The effect of heterogeneity in powders on the intensity of X-ray reflections," 640.
- , the influence of crystal size on the absorption factor as applied to Debye-Scherrer diffraction patterns, 215.
- Temperature coefficient of air-cored self-inductances, 693.
- , pressure and specific volume changes of a gas mixture under dissociation and re-association conditions, with combustion following the straight line law, $P=mV+n$, 680.
- Terminology and symbols in physics, 81.
- Thermal diffusion in mixtures of molecules of small mass difference, 30.
- endurance of glass rods, the effect of shape on the, 17.
- transpiration and reversible process for a degenerate gas, 760.
- Tolansky (S.), new contributions to interferometry. Part II.—New interference phenomena with Newton's rings, 120.

- Tolansky (S.), new contributions to interferometry. Part III.—The differential polarization phase change on reflection at a thin silver film, 179.
- , ———. Part IV.—The ghost images and scatter rings of the Fabry-Perot interferometer and their effects in hyperfine structure observations, 229.
- Tones of thin cylindrical tubes, on the, 571.
- Transient response in frequency modulation, 143.
- , letters and replies, 851, 852, 853.
- Tranter (C. J.) on a problem in heat conduction, 102.
- Trichromatic theory of colour vision, the development of the, 209.
- Two definitions of noise figure in radio receivers, a note on, 386.
- Two-dimensional compressible flow past a solid body in unlimited fluid or symmetrically placed in a channel, 549.
- Two methods of measuring ultra high frequency electric fields, 355.
- Ultrasonic diffraction grating, 192.
- Under-expansion in a steam-nozzle, with reference to an unusual case, 276.
- Upadhyaya (V. N.) and Gogate (D. V.), thermal transpiration and reversible process for a degenerate gas, 760.
- Viscosities and rheochors of aldehydes, nitriles and of secondary and tertiary amines, 619.
- Viscosities at the boiling point of some primary amines, cyclohexane and some of its derivatives, 57.
- Viscosities of unsaturated six-membered isocyclic compounds, 136.
- Viscous flow, solutions of some problems in, 250.
- Vivian (A. C.), an analysis of hardness, 765.
- Walker (W. J.), temperature, pressure and specific volume changes of a gas mixture under dissociation and re-association conditions, with combustion following the straight line law, $P=mV+n$, 680.
- Whittaker (E. T.), the new algebras and their significance for physics and philosophy, 1.
- Wight (K. C.) and Bratt (J. B.), the development of an automatic electric balance for research on aerodynamic stability, 588.
- Williams (G.), methods for the investigation of the properties of liquids at ultra high frequencies, 283.
- Wilson (W.), dimensions, 420.
- , kinematic relativity, 241.
- Winter (H. J. J.), the reception of Ohm's electrical researches by his contemporaries, 371.
- X-ray wave-lengths by the powder method: $\text{CrK}\beta_1$ and $\text{MnK}\beta_1$, the measurement of, 544.

END OF THE THIRTY-FIFTH VOLUME.

